

מדינת ישראל

משרד המשפטים

ייעוץ וחקיקה (משפט אזרחי)
ירושלים: כ"ה כסלו תשפ"ג
18 דצמבר 2022

חוות דעת: שימושים בתכנים מוגנים בזכויות יוצרים לצורך למידת מכונה

רקע

הולדתה של חוות דעת זו בפנייה מטעם המינהל למחקר, פיתוח אמצעי לחימה ותשתית טכנולוגית במשרד הבטחון (מפא"ת) למחלקת ייעוץ וחקיקה (אזרחי) בראשות המשנה ליועמ"שית עו"ד כרמית יוליס. הפניה עסקה בתכנית הדגל הממשלתית לתשתיות בינה מלאכותית, המקודמת בשיתוף מפא"ת, רשות החדשנות, ות"ת, משרד החדשנות, מדע וטכנולוגיה ומשרד האוצר. במסגרת תכנית זו, אמונה מפא"ת על פיתוח מודל לעיבוד שפה טבעית (NLP) בעברית ובערבית. תכנית זו נהגתה משזיהתה המדינה כשל שוק בתחום של עיבוד שפה טבעית בעברית ובערבית, הנובע בין היתר ממיעוט המשתמשים בשפות אלה ביחס לשפות אחרות ברמה הגלובלית. כשל זה מביא לפגיעה בכמות ובאיכות של מערכות מבוססות NLP בעברית ובערבית ביחס למקבילותיהן באנגלית. התכנית תיצר בעצמה מודל שפה מאומן שעל בסיסו יוכלו מיזמים אקדמיים, ממשלתיים ופרטיים לפתח שימושי NLP ספציפיים בעברית ובערבית.

השאלה שהובאה לפתחנו נגעה להיקף הזכות של מיזמים מבוססי בינה מלאכותית לעשות שימוש בתכנים המוגנים בזכויות יוצרים לצורך למידת מכונה. עם ראשית המחקר הסתבר כי אי הבהירות באשר ליחס בין דיני זכויות יוצרים ללמידת מכונה מהווה אתגר משפטי של ממש בישראל וברחבי העולם. לאור זאת הוחלט, בד בבד עם מתן ייעוץ נקודתי למפא"ת, לפרסם חוות דעת שתתמודד עם החסם של אי ודאות משפטית בשוק למידת המכונה ותאפשר שמירה על האיזונים הנאותים בדיני זכות יוצרים ועל המובילות הישראלית הן בתחום יצירת התוכן והן בתחום למידת המכונה.

לאחר כתיבת טיוטה ראשונית, קיימה מחלקת ייעוץ וחקיקה (אזרחי) שולחן עגול (במתכונת מקוונת) בשיתוף גורמים מהממשלה, מהאקדמיה ומהשוק הפרטי, בו הוצגו עיקרי חוות הדעת. בעקבות השולחן העגול נתקבלו עמדות כתובות מאת חברת גוגל ישראל, אגודת קומפוזיטורים, מחברים ומו"לים (אקו"ם) וחברת התמלוגים של יוצרי הקולנוע והטלוויזיה בישראל (תל"י). לצד הערות הציבור במהלך השולחן העגול, תגובות אלה נשקלו בכובד ראש ותרמו לעיצוב חוות הדעת.

חוות הדעת נכתבה על ידי עו"ד ד"ר ליטל הלמן, בליווי עו"ד הווארד פולינר, ראש אשכול קניין רוחני ובהנחיית עו"ד כרמית יוליס, המשנה ליועצת המשפטית לממשלה (משפט אזרחי). אנו מבקשים להודות ליועצת המשפטית לממשלה, עו"ד גלי בהרב-מיארה ולמנכ"ל משרד המשפטים, עו"ד ערן דויד, שהכירו את העבודה שנעשתה במחלקה בעניין זה ונתנו לה רוח גבית.

תודתנו נתונה על הערות מאירות עיניים מאת (בסדר אלפא-ביתי) עו"ד נעמי אברהם, ראש רשות הפטנטים עו"ד אופיר אלון, עו"ד זמר בלונדהיים, עו"ד עמרי בן-צבי, עו"ד ערן ברקת, עו"ד יוסף גדליהו, גב' הודיה גחלי-שוורץ, עו"ד טוני גרינמן, עו"ד תמר גנוניאן-פרקל, מר אלעד דביר, מר ערן דהן, מר דרור זמיר, מר ניר ינובסקי, עו"ד נועה מושיף, פרופ' מרים מרקוביץ-ביטון, פרופ' פיטר פיכט, פרופ' אורית פישמן-אפורי, מר ברק פלג, עו"ד איילת פלדמן, עו"ד גיא פרדיס, פרופ' גדעון פרחומובסקי, עו"ד סדריק יהודה צבע, עו"ד ד"ר אפי צמח, פרופ' משה קופל, ד"ר זיו קציר, עו"ד חיים רביה ועו"ד ד"ר יובל רויטמן. כן נבקש להודות על הערות מצויינות בעקבות הצגת חוות הדעת למחלקה האזרחית בייעוץ וחקיקה, למשתתפי השולחן העגול באוניברסיטת תל אביב בנושא עתיד הקניין הרוחני בצלן של טכנולוגיות חדשות, פרי ידיו של פרופ' אמיר חורי, למשתתפי הסמינר המחלקתי בקריה האקדמית אונו, למשתתפי השיח WIPO on IP and Frontier Technologies בזינבה ולמשתתפי המיט-אפ שקוים בנושא בניהולו של ד"ר זיו קציר, ראש התכנית הלאומית לתשתיות בינה מלאכותית. תודתנו נתונה גם למשנה ליועמ"ש (כלכלי) עו"ד מאיר לוי שמרכז את תחום הבינה המלאכותית במשרד המשפטים ולעו"ד צפרייר נוימן, היועץ המשפטי של רשות החדשנות, שהשותפות עמה תרמה רבות לעיצוב חוות הדעת. תודה מיוחדת לטל גבע ממפא"ת על הליווי המקצועי בתחום של למידת מכונה.

אנו מקווים כי חוות הדעת תאומץ במקרים שיידרשו ליישומה, וכי יהיה בכך כדי לקדם את שוק למידת המכונה תוך מתן תמריצים הולמים לקידום היצירה והתרבות בישראל.

עו"ד ד"ר ליטל הלמן
אשכול קניין רוחני

עו"ד הווארד פולינר
ראש אשכול קניין רוחני

עו"ד כרמית יוליס
המשנה ליועצת המשפטית לממשלה
(משפט אזרחי)

מדינת ישראל

משרד המשפטים

תקציר

חוות דעת זו נועדה להבהיר את היקף הזכות של מיזמים מבוססי בינה מלאכותית לעשות שימוש בתכנים המוגנים בזכויות יוצרים לצורך למידת מכונה. למידת מכונה (machine learning) היא היכולת של מחשב ללמוד בעצמו באופן אינדוקטיבי ממאגרי נתונים, והיא מהווה את התשתית הטכנולוגית למערכות של בינה מלאכותית. הבהרה של חוקיות השימוש בתכנים מוגנים לצורך למידת מכונה היא חשובה לנוכח מרכזיותן של מערכות בינה מלאכותית לכלכלה העולמית ולצורך שמירה על יתרונה התחרותי של ישראל כיצרנית של מערכות בינה מלאכותית.

כמפורט בחוות הדעת, איכותה של מערכת בינה מלאכותית תלויה ראש וראשית בהיקף הנתונים המוזנים לתוכה בשלב למידת המכונה, באיכותם ובגיוונם. נתונים כאמור יכולים לכלול סוגים מגוונים של תוכן, כגון תמונות, קבצי טקסט, קבצי קול, קבצי וידאו וכדומה, בהתאם למשימה שהמערכת לומדת לבצע. כאשר התכנים המשמשים לאימון המכונה אינם מוגנים בזכויות יוצרים, או שהם בבעלות מי שמאמן את המכונה, אין כל מניעה - ככל שבדיני זכויות יוצרים עסקינן - לעשות בהם שימוש. ואולם, הצורך בהיקף נרחב של תכנים איכותיים מחייב כמעט תמיד שימוש גם בתכנים המוגנים בזכויות יוצרים של גורמים נוספים. אי הבהירות באשר לנסיבות בהן ניתן לעשות שימוש בתכנים מוגנים במצבים אלה מהווה חסם משפטי לביצוע למידת מכונה אפקטיבית ולצמיחתו של שוק הבינה המלאכותית. אי וודאות זו מהווה אבן נגף גם בדרכם של יוצרים לאכוף את זכויותיהם בזירה זו.

על פי חוות דעת זו, מלבד מקרים חריגים, שימוש בתכנים מוגנים בזכויות יוצרים לצורך אימון מכונה חוסה תחת הסדרי השימושים המותרים בדיני זכויות היוצרים ולכן אינו מהווה הפרה של זכויות יוצרים. ראשית, כמפורט בחוות הדעת, שימוש בתכנים מוגנים לשם אימון מכונה יהווה לרוב שימוש הוגן. שנית, בחלק ממיזמי למידת המכונה, יחול סעיף 22 לחוק זכות יוצרים, העוסק בשימוש אגבי ביצירה. שלישית, ככל שהנתונים המשמשים ללמידה נמחקים בתום התהליך, יחול גם סעיף 26 לחוק זכות יוצרים בדבר יצירה ארעית. מסקנה זו מתיישבת גם עם המגמה הגלובלית בדין החל על הסוגיה של אימון מכונה.

לצד קביעה עקרונית זו, חוות הדעת כוללת גם סייגים. בפרט, אין חוות הדעת משתרעת על מקרים בהם מאגר המידע ללמידת המכונה אינו כולל מגוון רחב של תכנים אלא מורכב מיצירות של יוצרים בודדים, ומהווה תשתית למערכת בינה מלאכותית שתתחרה ביוצרים אלה בשווקים בהם הם פועלים. ככלל, אין חוות הדעת קובעת מסמרות באשר לתוצר של למידת המכונה. במצב בו התוצר שיפיק המחשב יהווה בעצמו הפרה של זכויות יוצרים, חוות הדעת לא תייצר נמל מבטחים לאחריות באופן קטגורי, ותחייב בחינה אד הוק של השימוש שעושה המיזם ביצירות המוגנות.

הסרת החסם של עמימות משפטית באמצעות חוות דעת זו נועדה לתמרץ פעילות רצויה בשוק למידת המכונה ולמקסם את התחרותיות של חברות ישראליות הן בתעשיית התוכן והן בתחום הבינה המלאכותית בשווקים גלובליים.

מדינת ישראל
משרד המשפטים

Abstract

This Opinion aims to shed light on the most fundamental question in the intersection between Machine Learning (ML) and copyright law: whether ML enterprises can make unauthorized use of copyrighted materials to train Artificial Intelligence (AI) systems. ML is the process that enables computers to learn autonomously from past data. ML thus provides the foundations for AI systems. The intersection between ML and copyright bears crucial importance. ML is becoming increasingly central to the global economy and to Israel in particular, owing to the leading position Israel holds as a producer of AI systems. Lifting the copyright uncertainties that surround this issue can spur innovation and maximize the competitiveness of Israeli-based enterprises in both ML and content creation.

As this Opinion explicates, the value of AI systems depends first and foremost on the quantity of the materials that the machine is fed at the ML stage, together with the diversity and quality of such materials. ML enterprises may use materials of various kinds—be it photos, text, sound, video, etc.—that are relevant for the task that the machine learns to eventually perform.

Obviously, copyright law creates no limitation on training machines on public domain materials or on works that are copyrighted or licensed by the said ML enterprise. Nevertheless, given the immense quantity of works that ML entails, the ML process will almost always require the use of works that are copyrighted by multitudes of third parties. Current Israeli copyright law does not provide clear guidance as to the infringing status of such a use. The vagueness of copyright law on this crucial issue yields severe legal uncertainty that may create a barrier for the development of the ML market and the growth of AI, without concomitantly yielding tangible benefits for copyright owners. This legal uncertainty may also form a hurdle for copyright enforcement in this arena.

This Opinion concludes that apart from certain circumstances, the use of copyrighted materials in the ML context is permitted under existing copyright doctrines. First, ML will typically be covered by the fair use doctrine. Second, some ML projects may fall under the doctrine that permits incidental uses of copyrighted materials. Third, in the (admittedly nonstandard) case where the copyrighted materials are erased at the end of the ML process, the enterprise may enjoy the protection provided under the doctrine of transient use. Notably, the conclusion of this Opinion—that training of ML systems is generally permitted under copyright law—is consistent with the approach of other legal systems around the globe.

While most ML cases come within the scope of permitted uses, this Opinion specifically excludes from its scope certain ML uses. For example, the safe harbor set forth in this Opinion would not apply to ML datasets that consists exclusively of works created by a single author in order to compete with this author in her existing markets. As a general matter, this Opinion does not apply to the *output* of the ML process. There may indeed be cases where the ML process would be protected under this Opinion, yet the output (or some of the outputs) of the resulting AI system would be deemed infringing. Overall, this Opinion endeavors to strike the right balance that will stimulate innovation in the ML domain while at the same time maintaining copyright integrity and enhance the incentive to create new works.

מדינת ישראל משרד המשפטים

א. מבוא

למידת מכונה מהווה את אחת המהפכות הדרמטיות של המאה ה-21. מהפכה זו מבטיחה תמורות משמעותיות בכל תחומי החברה והכלכלה, החל בשוק העבודה, עבור בתחומי התחבורה, הנדל"ן והאמנות וכלה בסוגיות של בטחון, רפואה ואינספור תחומים נוספים.¹ תרומתה הצפויה של למידת מכונה לכלכלה הגלובלית מוערכת בכ-15.7 טריליון דולר עד שנת 2030, ומדינות העולם מתחרות ביניהן על הנתח שיזכו לו מצמיחה אדירה זו.² בהתאם למעמדה כמובילה טכנולוגית בינלאומית, נמצאת מדינת ישראל בחוד החנית של פיתוחי למידת מכונה, הודות לחברות הזנק מקומיות, להשקעות מקומיות וזרות ולהקמת מרכזי פיתוח של חברות גלובליות בישראל.³

מטבע הדברים, החדשנות הגלומה בתחום של למידת מכונה מעוררת שאלות משפטיות שאין להן מענה ישיר בדין הקיים.⁴ חוות דעת זו עוסקת באחת משאלות אלה, ואולי במרכזית - או למצער - התשתיתית שבהן. כפי שיוסבר להלן, מערכות בינה מלאכותית באשר הן נשענות על מאגרי עתק של נתונים כבסיס ללמידה. דא עקא, מאגרי עתק של נתונים מורכבים תדיר מתכנים שנוצרו על ידי גורמים שונים, ודיני זכויות היוצרים מגבילים אפוא את השימוש בהם. השאלה המונחת ביסודה של חוות הדעת היא, אם כן, אחת השאלות הראשוניות של למידת מכונה בכל התחומים באשר הם: באיזו מידה מאפשרים דיני זכויות יוצרים לבינה מלאכותית ללמוד.

ברירת המחדל בדיני זכויות היוצרים היא כי העתקה של יצירה ושימושים אחרים בה טעונים הרשאה מאת בעל זכות היוצרים ביצירה.⁵ עם זאת, במספר מצבים, ברירת מחדל זו אינה מתקיימת. ראשית, לעתים נמצאת יצירה בנחלת הכלל, אם מכיוון שזכות היוצרים בה פקעה ואם מכיוון שמעולם לא היתה בה זכות יוצרים.⁶ שנית, לעתים מופצת יצירה בליווי רשיון "פתוח", לרוב Creative Commons (CC), המשנה את ברירת המחדל ומאפשר שימוש ביצירה לפי תנאי הרשיון.⁷ שלישית, והחשוב ביותר לענייננו, חוק זכות יוצרים עצמו מתיר שימושים מסויימים ביצירות

¹ ראו למשל אברהם טננבוים "קווי יסוד למשפט המכונית האוטונומית" **משפט מפתח** 3, 33, 36 (2015); Winnie Hu, *Driverless Cars Arrive in New York City*, N.Y. TIMES (Aug. 6, 2019), <https://www.nytimes.com/2019/08/06/nyregion/driverless-cars-new-york-city.html> (תחום התחבורה); Nicola Davis, *AI Equal with Human Experts in Medical Diagnosis, Study Finds*, GUARDIAN (Sept. 24, 2019), <https://www.theguardian.com/technology/2019/sep/24/ai-equal-with-human-experts-in-medical-diagnosis-study-finds> (תחום הרפואה); Jason Pontin, *How AI-Driven Insurance Could Reduce Gun Violence*, WIRED (Feb. 27, 2018), <https://www.wired.com/story/how-ai-driven-insurance-could-reduce-gun-violence/> (תחום הביטוח); Fortune Business Insights, *Home Automation Market to Expand at 12.1% CAGR and Reach USD 114 Billion by 2025*, GLOBENEWSWIRE (Apr. 27, 2021), <https://www.globenewswire.com/news-release/2021/04/27/2217431/0/en/Home-Automation-Market-Amir Mizroch, In to-Expand-at-12-1-CAGR-and-Rach-USD-114-Billion-by-2025.html> (תחום הנדל"ן); Israel, *a Standout Year for Artificial Intelligence Technologies*, FORBES, (MAR. 11, 2019), <https://www.forbes.com/sites/startupnationcentral/2019/03/11/in-israel-a-stand-out-year-for-artificial-intelligence-technologies/?sh=548db8ba30a8> (תחום הבטחון).

² *Sizing the prize: PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution*, PwC Global, <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-neil-savage-the-race-to-the-top-among-the-worlds-leaders-in-artificial-intelligence>, NATURE ; study.html INDEX, Sept. 9, 2020, <https://www.nature.com/articles/d41586-020-03409-8>

³ לירן ענתבי **בינה מלאכותית וביטחון לאומי בישראל** 85 (2020 INSS). ראו גם דאטאניישן "שנת שיא: סטארטאפים של בינה מלאכותית גייסו 25.2 מיליארד דולר ב-2018" **TheMarker** (21.3.19) <https://www.themarker.com/technation/datanation/1.7017976>

⁴ ע"א 9183/09 **The Football Association Premier League Ltd נ' פלוני**, פ"ד סה(3) 521, פס" 5 לפסק דינו של השופט מלצר (2012) ("ידוע הוא שהטכנולוגיה מקדימה לרוב את המשפט. במקרים אלה המחוקק ובתי המשפט נדרשים ליצוק את תמצית העקרונות הקיימים, הטובים והמבוססים - לקנקנים משפטיים חדשים משל אלה היו יין ישן המשתבח עם הזמן, וזקוק רק לכלי") (להלן: "עניין Premier League"); ע"ע 3782/12 **מפקד מחוז תל אביב-יפו במשטרת ישראל נ' איגוד האינטרנט הישראלי**, (פורסם בנבו, 2.3.2013), פסקה 23 לפסק דינו של השופט סולברג (2013) ("בידוע, כי המשפט מדדה בעצלתיים אחר חידושי העולם, וכי החקיקה אינה מדביקה את קצב התקדמות המדע וישומיו").

⁵ בעל זכות היוצרים הינו היוצר, מי שהיוצר העביר לו את זכות היוצרים או מי שהוא בעליה של זכות היוצרים על פי חוק. ראו ס' 11, 33-37 לחוק זכות יוצרים, התשס"ח-2007.

⁶ ראו ס' 4-5, 38 לחוק זכות יוצרים. ראו גם Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market, 14 September 2016, COM (2016) 593 final, 2016/0280 Council on copyright in the Digital Single Market, 14 September 2016, COM (Recital 8) ("[t]ext and data mining may also be carried out in relation to mere facts or data which are not protected by copyright and in such instances no authorisation would be required").

⁷ רשיונות CC משנים את ברירת המחדל מ"כל הזכויות שמורות" ל"חלק מן הזכויות שמורות", ומתירים שימוש ביצירה בתנאים, כגון מתן קרדיט ליוצר (CC-BY); הגבלת שימוש מסחרי (CC-NC) או הגבלת יצירה נגזרת (בכלל או בדרישה ל-"share alike"). ראו **About CC licenses Creative Commons**, <https://creativecommons.org/about/ccllicenses/>

מדינת ישראל משרד המשפטים

מוגנות ללא הרשאה או תמורה.⁸ אכן, ממושכלות היסוד של דיני זכויות יוצרים היא כי הגנה מוחלטת על יצירות אינה רצויה.⁹ דיני זכויות יוצרים שואפים לאזן בין האינטרס של יוצרים בהגבלת גישה ליצירות לבין האינטרס של הציבור (ושל יוצרים עתידיים) בהגברת גישה ליצירות.¹⁰ מטעם זה, זכויות היוצרים הן מעיקרא תחומות ומוגבלות, ויוצרות בד בבד זכויות בלעדיות ביצירה וסייגים לזכויות אלה.¹¹

דא עקא, היקף תחולתם של הסדרי השימושים המותרים על אימון מערכות לומדות לא הובהר עד עתה בישראל, והדבר הוא בעוכריהן של תעשיית הבינה המלאכותית ושל תעשיית התוכן כאחד. עמימות זו עלולה להוות חסם ליצירה של מאגרי נתונים איכותיים לצורך למידת מכונה ולהעמיס על מיזמים מבוססי בינה מלאכותית עלויות משפטיות באופן שיסכל את פעילותם או יקשה עליה.¹² בדומה, עמימות משפטית עשויה להקשות על יוצרים לאכוף את זכויותיהם אל מול משתמשים גם במקרים בהם הדבר ראוי ומוצדק. חוות דעת זו נועדה אפוא לפזר את הערפל סביב הסוגיה ולהבהיר במידת האפשר את ההסדרים החלים בה.

חוות דעת זו גורסת כי שלושה מהסדרי השימושים המותרים בחוק זכות יוצרים, התשס"ח-2007 (להלן: "החוק") רלוונטיים לבניית מאגרי נתונים ללמידת מכונה. ראש וראשית, דוקטרינת השימוש ההוגן הקבועה בסעיף 19 לחוק פורשת כנפיה על יצירת מאגרי נתונים במקרים התדירים של למידת מכונה. פרשנות זו מתיישבת עם לשון הסעיף, הולמת את תכליתו ועולה בקנה אחד גם עם הדין האמריקאי, ממנו ינק הדין הישראלי את הגנת השימוש ההוגן.¹³ שנית, במקרים מסויימים יוחל על פעילות זו סעיף 22 לחוק, העוסק ביצירה אגבית. זאת, בשל העובדה שמאגר הנתונים כולו, קל וחומר כל יצירה אינדיווידואלית הכלולה בו, אינם אלא אמצעי לאימון המחשב לבצע משימה אשר היא מטרתו של תהליך האימון. במצבים בהם למאגר הנתונים ערך כלכלי משמעותי משל עצמו וכן במצבים החורגים מלשון הסעיף, הגנה זו לא תחול.¹⁴ שלישית, במקרה (הלא תדיר) שהתכנים במאגר המידע נמחקים בתום תהליך האימון, ניתן להחיל על הסוגיה את ההסדר הקבוע בסעיף 26, העוסק בהעתקה ארעית. הגם כי מדובר בפרשנות מרחיבה למושג העתקה ארעית, תואמת היא הן את לשון החוק והן את הפרשנות שניתנה למושג זה בדין האירופי.¹⁵

עיקרו של דבר, על פי חוות דעת זו, במרבית המקרים דיני זכויות יוצרים מתירים הכללתם של תכנים מוגנים במאגרי נתונים ללמידת מכונה. תפיסה זו נתמכת בפרשנות תכליתית לדיני זכות יוצרים, וזאת משני טעמים עיקריים. הטעם הראשון נעוץ באנאלוגיה ללמידה אנושית אינדוקטיבית. אכן, כשבלמידה אנושית עסקינן, בסיס המידע המאפשר למידה "נמצא" במוחו של

⁸ סעיפים 19-30 ו-32 לחוק זכות יוצרים; ס' 33 לפקודת זכות יוצרים, 1911. ראו גם להלן פרק ג.
⁹ ראו למשל ע"א 8393/96 מפעל הפיס נ' The Roy Export Establishment Company, פ"ד נד(1) 577, 596 (2000) ("יבואנו להגן על היצירה המקורית יש לתת את הדעת גם על כך שהגנה מוגברת יתר על המידה עלולה לבלום התפתחות תרבותית וחברתית, הסומכת על הישגי העבר.") (להלן: "עניין מפעל הפיס"). ראו גם: Abraham Bell & Gideon, *Reinventing Copyright and Patent*, MICH. L. REV., Vol. 113, P. 231 (2014).
¹⁰ ראו דברי ההסבר לחוק זכות יוצרים, חלק כללי, התשס"ח-2007, ה"ח הממשלה 1116: "דיני זכות יוצרים נועדו לקבוע הסדר שמטרתו הגנה על יצירות, תוך איזון בין אינטרסים שונים, לטובת הציבור"; 5097/11 טלראן תקשורת (1986) בע"מ נ' צ'רלטון בע"מ, פס' 13 לפסק-דינו של השופט זילברטל (נבו) 2.9.2013. (מגדיר את דיני זכויות יוצרים כפרי של "איזונים בין האינטרסים של יוצרים עתידיים ושל משתמשים שונים במרחב תמרון יצירתי הולם.") (להלן: "עניין טלראן"); ע"א 326/00 עיריית חולון נ' אן.אם.סי. מוסיקה בע"מ פ"ד נז(3) 658, 663-664 (2003) ("והנה – במעגל הפרשני נתקלים אנו בדיני זכויות היוצרים במתח בין אינטרסים שונים והם, בעיקר, בין האינטרס הקנייני של הטוען לזכות לבין אינטרס הכללי"). ר' גם אורית פישמן-אפורי, "זכויות תרבותיות וזכויות אדם: כלי לפיתוח מאוזן של דיני זכויות יוצרים בישראל" משפטים 499 (2007).

¹¹ הסייגים בחוק זכות יוצרים כוללים, בין היתר, הגבלה של תקופת זכויות היוצרים, אי-החלת זכויות יוצרים על רעיונות, שיטות ונתונים ורשימה של שימושים מותרים ביצירה ללא רישיון וללא תשלום בגין השימוש. ראו למשל ס' 5 וכן פרק ד לחוק זכות יוצרים. ראו גם טוני גרינמן, **זכות יוצרים** מהדורה שניה, כרך 4 א (2008) (להלן: גרינמן); מיכאל בירנהק, "יצירה משפטית: שימוש הוגן בדיני זכויות יוצרים", **משפט, תרבות וספר - ספר נילי כהן 1**, 1 (ע. גרוסקופף ופ. לביא עורכים) (2017) ("דיני זכויות יוצרים מבקשים לעודד את היצירה על ידי יצירת כלי משפטי, שנועד לאפשר פעולה עצמית בשוק חופשי, מעין זכות קניין, אך כזו שהיא מעוצבת מראש, מתוחמת ומוגבלת").

¹² חשש זה אינו ייחודי לישראל. ראו למשל European Commission (2016), Commission Staff Working Document, Impact Assessment on the modernisation of EU copyright rules, 14 September 2016, SWD (2016) 301 final, Part 3/3, p. 94 ("researchers are faced with legal uncertainty with regard to whether and under which conditions they can carry out TDM on content they have lawful access to").

¹³ ראו 17 U.S.C. § 107. ראו גם להלן פרק ד. להשוואה בין שימוש הוגן בדין האמריקאי והישראלי, ראו ניל נתנאל "שימוש הוגן ישראלי מנקודת מבט אמריקנית" **יוצרים זכויות: קריאות בחוק זכות יוצרים** 377 (מיכאל בירנהק וגיא פסח עורכים) (2009).

¹⁴ ראו להלן פרק ג.2.ב.

¹⁵ ראו להלן פרק ג.2.ג.

מדינת ישראל משרד המשפטים

האדם הלמד מבלי שתידרש לכך הגנה בדיני זכויות יוצרים.¹⁶ במובחן מבן אנוש, מחשב אינו יכול - כיום - ללמוד מ"קריאת" תכנים מבלי שאלה יועתקו קודם לכן למאגר נתונים.¹⁷ הגנה בזכויות יוצרים נדרשת בהקשר של למידת מכונה רק בעטייה של מגבלה טכנית-טכנולוגית זו (שעשויה גם להשתנות עם התקדמות הטכנולוגיה). מימוש ההבטחה הגלומה בבניה מלאכותית, לאפשר למכונות להתחקות אחר הלמידה האנושית כדי לבצע משימות מקבילות באופן איכותי, מותנה בכך שלמידה תתאפשר עבור מכונות הגם שטכנית למידה על ידי מכונה כוללת ייצור של העתק. לא נסתר מעינינו כי אנלוגיה ישירה בין למידת מחשב ללמידה אנושית היא מוגבלת, שהרי למידת המחשב אין מטרתה הרחבת דעת בן אנוש, אלא קידום אינטרסים כלכליים (בדרך כלל) של הגוף המפעיל את המערכת. ואולם לטעמנו הליך הלמידה ככזה ראוי שיזכה להגנה הן כשהלומד הוא אנושי והן כשאין הוא אלא מכונה, כיוון שבשני המקרים, הלמידה מהווה תשתית לפיתוח יכולות חדשות ומעודדת התפתחות וקידמה. במצבים בהם דיני זכויות יוצרים מצדיקים חסימה של תוצרי למידה מסויימים, ראוי שהחסימה תעשה בשלב הפקת התוצר ולא בשלב הלמידה, שיכול להוות תשתית גם לתוצרים לגיטימיים.

שנית, החלת הסדרי השימושים המותרים על מרבית המקרים של למידת מכונה מאזנת היטב בין האינטרסים של היוצרים לבין אלה של הציבור, איזון שהוא כאמור לחם חוקו של משטר זכויות היוצרים מימים ימימה ונשמת אפו.¹⁸ מחד גיסא, החלת הסדרי השימושים המותרים על למידת מכונה תאפשר לבניה מלאכותית להתפתח ללא הגבלה מכוח דיני זכויות יוצרים באשר לבסיס המידע שיכול לשמש ללמידת המכונה. ללא החלתם של הסדרי השימושים המותרים, כשלי שוק בתחום של גישה דיגיטלית ליצירות צפויים לסכל יצירה של מאגרי מידע אפקטיביים. כך, כפי שידון להלן, עלויות העסקה הכרוכות בתהליכי האיתור וקבלת הרשיונות מאת בעלי זכויות היוצרים, ויכולתו של כל אחד ממליוני בעלי זכויות לעכב את הפרוייקט ("בעיית הסחטן"), יטילו אפקט מצנן משמעותי על יזמות בתחום של למידת מכונה ועלולות להפוך את הפעילות של מיזמי בניה מלאכותית לבלתי כדאית או לבלתי מעשית.¹⁹

מאידך גיסא, כפי שידון להלן, הפגיעה בבעלי זכויות יוצרים בגין החלת הסדרי השימושים המותרים על מאגרי נתונים ללמידת מכונה היא זניחה, אם בכלל קיימת. שימוש זה בתכנים אינו פוגע כמלוא הנימה בשווקים קיימים של בעלי זכויות יוצרים, ובמרבית המקרים אינו שולל מהם רווח פוטנציאלי.²⁰ יושם אל לב כי ממילא, כל יצירה אינדיווידואלית מהווה רכיב יחיד במאגר נתונים עצום מימדים אשר לאף יצירה אין בו משקל סגולי. ערך השוק של רשיון השימוש ביצירות ישקף עובדה זו ויהיה ממילא מזערי, וכנראה נמוך משמעותית מהעלויות הכרוכות בהשגת רשיון שימוש בכל אחת מהיצירות. ודוק. הצורך לעצב מבנים קנייניים באופן דינאמי כדי להגשים תכליות חברתיות מעוגן היטב בפסיקה ובספרות.²¹ צורך זה מובהק הוא בהקשר של זכויות יוצרים, שעל אף מעמדן הקנייני אינן זהות לקניין בנכס פיסי, וכוללות מיניה וביה הגבלות וסייגים.²² אכן, גם אם נאמר כי זכויות יוצרים מהוות קניין במובנו הרחב, עדיין יש מקום לבחון את השתרעותו של קניין זה והזכויות המוגנות בו מצב בו ניצב בפנינו אינטרס חברתי, ציבורי, כלכלי, כבד משקל, של קידום טכנולוגיות מתקדמות, הטמעתן בשוק, האצת הקידמה והבאת בשורת התפתחות. לטעמנו,

¹⁶ ראו למשל Thomas Margoni, *Artificial Intelligence, Machine Learning and EU Copyright Law: Who Owns AI?*, CREATE Working Paper (Dec. 2018), available at SSRN: <https://ssrn.com/abstract=3299523> ("when humans learn a new language they usually store the training information (e.g. the text book used to learn it) as an electrochemical trace in the area of the brain dedicated to language. Humans do not need a copyright exception in order to store that copy" (להלן: "Margoni").

¹⁷ שם, בעמ' 4 ("In NLP, as well as in most text analytic fields, algorithms "learn" abstract probabilistic models from texts annotated with labels... in order to predict such labels on unseen text. They do this by storing the relevant information in a separate file, the "trained model").

¹⁸ ראו לעיל ה"ש 9-10. ראו גם Universal Declaration of Human Rights, GA Res 217 (III) A, UN Doc A/RES/217(III) art 27 (10 December 1948) (מכירה בפסקה הראשונה בזכות הגישה למידע ובפסקה השניה בזכות היוצר להגבלת גישה ליצירותיו); JAMES BOYLE, SHAMANS, SOFTWARE & SPLEENS: LAW AND THE INFORMATION SOCIETY 38 (1996); ROBERT P. MERGES, JUSTIFYING CONSTRUCTION OF THE INTELLECTUAL PROPERTY 136 (2011).

¹⁹ ראו להלן ה"ש 87-91 וטקסט נלווה.

²⁰ ראו עוד להלן פרק 2.א.

²¹ ראו למשל רע"א 6339/97 **רוקר נ' סלומון**, פ"ד נה(1)199, 280 ("הקניין נועד לבטא את שליטתו של האדם ואת אישיותו. עם זאת, לקניין היבט ציבורי... והשימוש בו צריך לשרת את אינטרס הציבורי"); ת"א (מחוזי ת"א) 2177-05 **ADIDAS-SALMON נ' יאסין** (פורסם בנבו, 13.12.2010), בעמ' 30 ("יש לעצב זכויות קניין באופן שמייצג את ההתפתחות החברתית והתרבותית הראויה."); חנוך דגן "קריאה קניינית: המוסד הקנייני המתחדש של זכות יוצרים" **יוצרים זכויות - קריאות בחוק זכות יוצרים** (תשס"ט) 47 ("שיח קנייני כזה הוא מטבעו תכליתי ודינמי: הוא מאפשר – ואולי אפילו מזמין – פיתוח תמידי (אם גם זהיר) של מוסדות הקניין השונים באופן שישפר את פועלם בקידום אותן תכליות.");

²² ראו לעיל ה"ש 11.

מדינת ישראל

משרד המשפטים

מקומות בהם זכויות פרט ניצבות אל מול אינטרסים ציבוריים משמעותיים, יש ליתן את הדעת גם למאזן ביניהן על מנת להשיא את התועלת המצרפית המיטבית. הניתוח בחוות דעת זו, לפיו זכויות היוצרים אינן משתרעות במקרה הרגיל על יצירת מאגרי נתונים ללמידת מכונה, משקפת נאמנה, לדעתנו, את האיזון הפנימי בדיני זכויות היוצרים.

לצד אמירות אלו, וכפי שעולה מהן, חוות הדעת מעניקה הגנה לתהליך של למידת מכונה במצבים בהם אכן מדובר במאגרי נתונים עצומי מימדים שכל יצירה מהווה רכיב זניח ובלתי מהותי בהם. מצבים בהם מאגרי הנתונים הם מצומצמים מאד ומכוונים לכלול יצירות מסוג מסויים (למשל כדי לייצר תוצרים זהים), עשויים להיות מוחרגים מחוות הדעת כמפורט להלן.

היקפה של חוות דעת זו ומגבלותיה ידונו בפירוט להלן.²³ יובהר כבר עתה כי חוות דעת זו אינה חלה על **התוצר** של מערכות בינה מלאכותית שהתבססו על למידת מכונה אלא רק על תהליך הלמידה עצמו. משמעות הדבר היא כי השאלה אם **התוצר** מפר זכויות יוצרים תבחן אד הוק לפי הכללים הרגילים של דיני זכויות יוצרים וחוות דעת זו לא מקנה לו הגנה אפרוירית.²⁴ כן יצויין כי חוות דעת זו אינה משתרעת על הגבלות לשימוש בתכנים שמקורם בדינים אחרים לבר מזכויות יוצרים, כגון דיני הפרטיות, חסיונות, בטחון המדינה, אתיקה וכיוצאים באלה.²⁵ לבסוף, יודגש כי חוות דעת זו חלה על הדין הישראלי בלבד, ואין בכוחה לחלוש על הסוגיה מחוץ לגבולות מדינת ישראל או כאשר הדין החל על הסוגיה אינו הדין הישראלי.²⁶

הדיון להלן יתקיים כדלקמן. הפרק השני יציע רקע טכנולוגי לדיון במאגרי נתונים ללמידת מכונה. הפרק השלישי יפרט את המסגרת הנורמטיבית של דיני זכויות יוצרים הנדרשת לביורר הסוגיה, על ברירות המחדל שבה והסדרי השימושים המותרים, ויישם אותה על הסוגיה של למידת מכונה. הפרק הרביעי יעסוק במשפט משווה ויפרט כיצד מדינות אחרות התמודדו עם הסוגיה. הפרק החמישי יעסוק בהיקפה של חוות הדעת וידונו בו גם מקרים אשר אינם חוסים תחתיה. במסגרת הסיכום, בפרק השישי, תדון בין היתר האפשרות לקדם תיקון חקיקה (או חקיקת משנה) ברוח האמור בחוות הדעת, ותובע העמדה כי מוטב להתבסס על הסדרי השימושים המותרים בדין הקיים, לפחות לעת עתה.

ב. רקע: למידת מכונה ומערכות מבוססות בינה מלאכותית

למידת מכונה (machine learning) מהווה את התשתית הטכנולוגית למערכות בינה מלאכותית.²⁷ למידת מכונה היא היכולת של מחשב ללמוד בעצמו באופן אינדוקטיבי ממאגר נתונים. במובחן ממערכות "לא לומדות", בהן מתכנת מזין בעצמו למחשב את המידע הנדרש לפעולת התוכנה, במערכות המבוססות על למידת מכונה, המעורבות האנושית מתמצה בכתיבת האלגוריתמים שמנחים את למידת המחשב וביצירת מאגר נתונים שיהווה בסיס ללמידה. לאחר שהמחשב "מתאמן" על מאגר הנתונים שנבנה עבורו, הוא יכול ליישם את מסקנותיו גם על נתונים חדשים שלא הוצגו לו קודם לכן.²⁸ נתונים חדשים אלה יכולים בהמשך להיות מוזנים למאגר המידע ולהפוך גם הם לבסיס ללמידה. כתוצאה מכך, ביצועיו של המחשב משתפרים באופן עקבי.²⁹

²³ ראו להלן פרק ד.

²⁴ ראו להלן פרק ד.

²⁵ ראו למשל (2016) 1143, 1165 COLUM. L. Rev. 1143, Orin Kerr, *Norms of Computer Trespass*.

²⁶ כללי ברירת הדין בדיני זכויות יוצרים הם מורכבים וחורגים מעניינה של חוות דעת זו. ראו למשל ע"א 2790-93 Eisenman נ' קימרון, פ"ד נד(3) 817 (2000).

²⁷ המושג למידת מכונה (Machine Learning) נטבע לראשונה כבר ב-1959. ראו 1959, Arthur L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers. I*, in *Computer Games I* 335 (David N.L. Levy ed., 1988) אולם הזינוק הטכנולוגי שאפשר את הדבר מתוארך ל-2010 בערך. ר' גם Michael Veale & Irina Brass, *Administration by Algorithm? Public Management Meets Public Sector Machine Learning*, in ALGORITHMIC REGULATION 121, 125 (Karen Yeung & Michael Lodge eds., 2019) ("a machine learns when its performance at a certain task improves with experience.")

²⁸ Will Knight, *This AI-Generated Musak Shows Us The Limit Of Artificial Creativity*, MIT <https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart>; Ayush Pant, *Introduction to Machine Learning for Beginners*, TOWARD DATA SCIENCE (Jan. 7, 2019) <https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-eed6024fdb08>

²⁹ ש.ס.

מדינת ישראל

משרד המשפטים

על מנת לאמן מערכת לומדת נדרש מאגר נתונים רחב היקף, רלוונטי ומגוון.³⁰ הנתונים במאגר עשויים לכלול אחד או יותר מסוגי מידע שונים - תמונות, טקסטים, קבצי קול, קבצי מולטימדיה וכיוצא באלה, בהתאם לאופי המשימה שנועדה המערכת לבצע. גיבוש מאגר הנתונים נעשה בשני שלבים. השלב הראשון כולל איסוף מידע - לרוב באמצעות תוכנות אוטומטיות ("בוטים"). בשלב השני מתבצע "ניקוי" של המידע שנאסף ממידע בלתי רלוונטי שנאסף איתו, לרבות עיצוב, קישורים וכיוצא באלה.³¹ שני שלבים אלה הם עתירי זמן ומשאבים ותובעים את החלק הארי של התשומות הכרוכות במיזמים של למידת מכונה.³²

בתום שלבים אלה, אלגוריתמים של למידה שמופעלים על הנתונים מייצרים מודל סטטיסטי, שיכול לעמוד בפני עצמו או להוות 'שכבת בסיס' לפיתוח של משימות קצה ספציפיות. כך למשל, מערכת לעיבוד שפה טבעית (NLP - Natural Language Processing) תתאמן על מאגר נתונים רחב של תכנים טקסטואליים. בסוף תהליך האימון, יוצר מודל שיוכל "להבין" שפה טבעית באופן כללי. משימות קצה ספציפיות יוכלו לעשות שימוש במודל כדי ליצור מערכות מגוונות מבוססות-NLP, כגון מערכת לסיכום אוטומטי של מסמכים, מערכת זיהוי קולי, מערכת לשירות לקוחות אוטומטי ושימושים רבים נוספים.

משימות קצה ספציפיות טעונות ביצוע של שלב נוסף - 'תיוג' של מאגר המידע, דהיינו הזנה למערכת של דוגמאות לאופן בו ניתן ללמוד מן המידע. לשם המחשה, מערכת שנועדה לזהות בעלי חיים בתמונות תוכל להתאמן על מידע המאורגן כטבלה בת שתי עמודות: בראשונה ימצאו התמונות עצמן ובשניה, ביאור של החיה שבתמונה. עיבוד זה של הנתונים הגולמיים נקרא 'תיוג', והוא נעשה לרוב על ידי צוותים אנושיים או על ידי שילוב של אדם ומכונה.³³ המערכת תלמד מהמאגר המתויג להבחין בין בעלי חיים שונים, ובתום תהליך האימון, תוכל לזהות חיות גם בתמונות שלא נחשפה אליהן בעבר.

עוד יובהר כי מאגר נתונים אינו אובייקט סטטי אלא יצירה דינאמית המתפתחת לאורך חיי המיזם.³⁴ אכן, מאגר המידע מתעדכן כאמור באמצעות הוספת נתונים חדשים לאחר שהמערכת כבר פועלת. מאגרי מידע מתעדכנים גם למטרות אחרות, כגון תיקון טעויות, צורך (משפטי או אחר) בהסרת פריטים מהמאגר, שינוי בשימושי המערכת (use-cases), וכיו"ב. מאגר הנתונים מהווה אפוא יסוד תשתיתי וקרדינלי למערכות בינה מלאכותית לאורך כל חיי המיזם.³⁵

כפי שנובע מהמתואר לעיל, רמתה של מערכת בינה מלאכותית תלויה ראש וראשית בהיקף הנתונים המוזנים למערכת בשלב למידת המכונה, באיכותם ובגיוונם.³⁶ ישנו איפוא הכרח טכנולוגי ומערכתי לעשות שימוש במאגר נתונים רחב ורלוונטי ככל הניתן על מנת לטייב את ביצועי המערכת. דא עקא, דיני זכויות יוצרים מגבילים את הזכות לעשות שימוש בתכנים שנוצרו על ידי אדם אחר.³⁷

³⁰ Google, *Training and Test Sets: Splitting Data*, <https://developers.google.com/machine-learning/crash-course/training-and-test-sets/splitting-data>, ראו למשל

³¹ Jiawei Han, Micheline Kamber and Jian Pei (2011), *Data Mining: Concept and Techniques*, 3rd. ed., Waltham, USA: Morgan Kaufmann

³² ראו למשל רשות החדשנות, קול קורא - הקמת מאגרי מידע בעברית ו/או בערבית מדוברת, https://innovationisrael.org.il/kol-kore/Infrastructure_databases (נצפה לאחרונה: 8 באפריל, 2022) (קול קורא המתקבץ יצירה של מאגר מידע למיזם NLP - הכולל איסוף ו"ניקוי" של מידע בעד 550,000 ש"ח לכל יישום).

³³ Margoni, *לעיל* ה"ש 16, בפסקה II. למידה ממידע שאינו מתייג מכונה unsupervised learning.

³⁴ ראו למשל 1.4 MIMIC-III, <https://physionet.org/content/mimiciii/1.4> (נצפה לאחרונה: 5.2.22) (מתאר פרוייקט עצום מימדים למאגר נתונים של מידע רפואי המתוחזק ומתעדכן באופן תדיר על ידי גורמי מחקר ותעשייה).

³⁵ ראו גם Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579 (2018) (דנה בזכות יוצרים במאגרי המידע עצמם, נושא החורג מחוות דעת זו).

³⁶ המונח המקובל בתעשייה להגדרת התופעה הוא "Garbage in garbage out". להרחבה ראו Garbage in, garbage out, WORLD WIDE WORDS, <http://www.worldwidewords.org/qa/qa-gar1.htm>. ראו למשל מחקרים המצביעים על הטיות חריפות בפעולת מערכות של בינה מלאכותית על בסיס מאגרי מידע מצומצמים או לא מגוונים

דיים: SARA WACHTER-BOETTCHER, *TECHNICALLY WRONG: SEXIST APPS, BIASED ALGORITHMS, AND OTHER THREATS OF TOXIC TECH* (2017); Wenying Wu et al., *Gender Classification and Bias Mitigation in Facial Images*, in 12TH ACM CON. WEB SCIENCE 106 (2020), <https://doi.org/10.1145/3394231.3397900>;

Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579 (2018).

³⁷ שאלת ההגנה על יצירות שנוצרו על ידי יוצרים שאינם אנושיים נדונה לאורך השנים ברחבי העולם. ראו למשל *Naruto v. Slater*, No. 16-15469 (9th Cir. 2018) (עוסק בתמונה שצולמה על ידי שימפנזה). באחד המהלכים האחרונים בתחום, סרבה ספריית הקונגרס בארה"ב להכיר בזכות יוצרים ביצירה שנוצרה על ידי מחשב. ראו

Copyright Office, *Second Request for Reconsideration for Refusal to Register A Recent Entrance to Paradise*, (Correspondence ID 1-3ZPC6C3; SR # 1-7100387071), Feb. 14, 2022

<https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf>. בסוף

מדינת ישראל

משרד המשפטים

האם בכוחם של דיני זכויות יוצרים למנוע הכללתם של תכנים במאגר נתונים ללמידת מכונה? לשאלה זו נפנה עתה.

ג. המסגרת הנורמטיבית - שימוש בחומרים המוגנים בזכות יוצרים

להלן תפורט המסגרת הנורמטיבית לדיון בשימוש ביצירות המוגנות בזכות יוצרים. כפי שנראה להלן, נקודת המוצא במשפט הישראלי היא כי יצירות מוגנות ככלל בזכות יוצרים למשך חיי היוצר ושבעים שנה לאחר מותו.³⁸ יצירות מוגנות מקנות לבעליהן זכות בלעדית לביצוע שימושים שונים ביצירה, ומשמעות הדבר היא שלא ניתן לעשות שימוש משימושים אלה ביצירות מוגנות אלא ברשותו של בעל זכות היוצרים. על אף האמור, לצד משטר הזכויות האקסקלוסיביות, מתיר החוק עשיית שימושים ביצירות בתנאים מסויימים ללא הרשאה מאת היוצר וללא מתן תמורה בגין השימוש.³⁹ חלקו הראשון של פרק זה יבחן אימתי למידת מכונה עשויה לחשוף את מבצעה לטענות בדבר הפרת זכויות יוצרים. בחלקו השני של הפרק יידונו הסדרי השימושים המותרים שעשויים לחול על למידת מכונה.

1. הגנת זכויות יוצרים

המסגרת הנורמטיבית לבחינת חוקיות השימוש בתכנים מוגנים היא חוק זכות יוצרים, התשס"ח-2007. חוק זה כולל הוראות העוסקות בהיקף ההגנה על יצירות ותואם גם להוראות האמנות הבינלאומיות שישראל חתומה עליהן. נקודת המוצא היא כי סף ההגנה על יצירות הוא נמוך, דהיינו, לא נדרש הרבה על מנת שיצירה תזכה בהגנת זכות יוצרים. כך, סעיף 4(א) לחוק קובע כי -

”זכות יוצרים תהא ביצירות אלה:

- (1) יצירה מקורית שהיא יצירה ספרותית, יצירה אמנותית, יצירה דרמטית או יצירה מוסיקלית, המקובעת בצורה כלשהי;
 - (2) תקליט;
- ובלבד שהתקיים לגבי היצירות האמורות אחד התנאים הקבועים בסעיף 8 או שיש בהן זכות יוצרים מכוח צו לפי סעיף 9.”

מבלי להכנס לפרטי התנאים המפורטים בחוק, הסף הנמוך של הגנת זכויות יוצרים משמעו כי איסוף רחב היקף של תכנים, כגון זה הנדרש עבור אימון מכונה, יעלה בחכתו כנראה גם תכנים מוגנים בזכויות יוצרים. הוסף לסף ההגנה הנמוך את העובדה כי זכויות יוצרים נותרות בתוקפן שבעים שנה אחרי מות היוצר, והרי לך הכרה כי כמעט כל מאגר מידע שיבנה לצורך למידת מכונה יכלול תכנים מוגנים לרוב. ודוק. יצירות מוגנות אינן בהכרח מסומנות ככאלה ואינן רשומות במרשם כלשהו.⁴⁰ אשר על כן, אין בידם של משתמשים פוטנציאליים ביצירות להבחין בנקל בין יצירות מוגנות לשאינן מוגנות בטרם יוחל במלאכת האיסוף וההעתקה של החומרים. מטעמים

עומת זאת נפסק לטובת מתן הגנה כאמור. ראו Paul Sawers, *Chinese Court Rules AI-Written Article is Protected by Copyright*, VENTUREBEAT (Jan. 10, 2020), <https://venturebeat.com/2020/01/10/chinese-court-rules-ai-written-article-is-protected-by-copyright/>. למשל Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. PITT. L. REV. 1185 (1986); Clark D. Asay, *Independent Creation in a World of AI*, 14 FLA. INT'L. U. L. REV. 201 (2020); Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 5, 2; Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, 34 BERKELEY TECH. L.J. 343 (2019); James Grimmelmann, *There's No Such Thing as a Computer-Authored Work—And It's Liubov Maidanyk, Artificial Intelligence and Sui ; a Good Thing, Too*, 39 COLUM. J. L. & ARTS 403 (2016) *Generis Right: A Perspective for Copyright of Ukraine?*, 3(11) ACCESS TO JUSTICE IN EASTERN EUROPE 144–154 (2021).

³⁸ ישנן תקופות הגנה מיוחדות, כגון לעניין תקליט, גופן ויצירות של המדינה. ראו ס' 38-44 לחוק זכות יוצרים.

³⁹ ראו פרק ד לחוק זכות יוצרים.

⁴⁰ ראו ס' 25(2) לאמנת ברן להגנת יצירות ספרותיות ואמנותיות, כ"א 21, 581 (נפתחה לחתימה ב-1949) (אוסר על התניית זכות היוצרים בתנאים פורמליים כגון רישום). בארצות הברית, ישנו רישום של זכויות יוצרים אך הרישום אינו מהווה תנאי לתוקף זכות היוצרים. ראו “[R]egistration is not a condition of copyright protection.” 17 U.S.C. § 408(a) (2018) (“[R]egistration is not a condition of copyright protection.”); *Kernel Records Oy v. Mosley*, 694 F.3d 1294, 1301 (11th Cir. 2012); *Automation by Design, Inc. v. Raybestos Prods. Co.*, 463 F.3d 749, 752 n.1 (7th Cir. 2006).

מדינת ישראל משרד המשפטים

אלה, על מיזמי בינה מלאכותית להניח כי לפחות חלק מהיצירות בהן עושים הם שימוש במאגר הנתונים, וכנראה מרביתן, מוגנות הן בזכויות היוצרים.

יצירות מוגנות זכאיות הן לשתי מערכות של זכויות: זכות יוצרים לפי ס' 11 לחוק (המכונה: "הזכות הכלכלית") וזכות מוסרית לפי סעיף 46 לחוק. להלן יפורטו הסדרים אלה.

א. הזכות הכלכלית

הכלל החל על יצירות מוגנות הוא שלבעל זכות היוצרים הזכות הבלעדית לעשות ביצירות שימוש מן השימושים המוגדרים בחוק ("אגד הזכויות"). משמעות הדבר היא כי אלא אם חל אחד מהסדרי השימושים המותרים אשר יידונו להלן, אין לבצע את הפעולות המנויות בסעיף 11 ללא אישורו של בעל זכות היוצרים. וזוהי לשון הסעיף:

"זכות יוצרים ביצירה היא הזכות הבלעדית לעשות ביצירה, או בחלק מהותי ממנה, פעולה, אחת או יותר, כמפורט להלן, בהתאם לסוג היצירה:

- (1) העתקה כאמור בסעיף 12 - לגבי כל סוגי היצירות;
- (2) פרסום - לגבי יצירה שלא פורסמה;
- (3) ביצוע פומבי כאמור בסעיף 13 - לגבי יצירה ספרותית, יצירה דרמטית, יצירה מוסיקלית ותקליט;
- (4) שידור כאמור בסעיף 14 - לגבי כל סוגי היצירות;
- (5) העמדת היצירה לרשות הציבור כאמור בסעיף 15 - לגבי כל סוגי היצירות;
- (6) עשיית יצירה נגזרת כאמור בסעיף 16, ועשיית הפעולות המנויות בפסקאות (1) עד (5) ביצירה הנגזרת כאמור - לגבי יצירה ספרותית, יצירה אמנותית שאינה גופן, יצירה דרמטית ויצירה מוסיקלית;
- (7) השכרה כאמור בסעיף 17 - לגבי תקליט, יצירה קולנועית ותוכנת מחשב."

טכנית, יצירת מאגר נתונים הכולל יצירות מוגנות משמעו העתקת התכנים (זכות המעוגנת בס' 11(1) לחוק), שמירתם במאגר הנתונים (טכנית, העתקה נוספת של התכנים), ולעיתים גם ביצוע שינויים בהם. מעבר לשינויים מינוריים ביצירות, מיזמים של למידת מכונה מבצעים לעתים מזומנות גם מניפולציות על היצירות הכלולות במאגר המידע ומייצרים מהן יצירות נגזרות על מנת להרחיב את היקף מאגר המידע ואת גיוונו. פרקטיקה זו עשויה להוות עשיית יצירה נגזרת והעתקתה, כאמור בס' 11(6). אשר על כן, בהעדר תחולה להסדרי השימושים המותרים, מאגר נתונים לצורך אימון מכונה עשוי לחשוף את יוצרו לטענות בדבר מספר הפרות של זכות היוצרים, וזאת מכל אחד מבעלי הזכויות ביצירות הרבות הכלולות במאגר הנתונים.

ב. הזכות המוסרית

לצד הזכויות הכלכליות, מעניק החוק ליוצר היצירה זכות מוסרית בעלת שני היבטים, הקבועה בס' 46 לחוק זכויות היוצרים בהאי לי שנא -

"זכות מוסרית ביחס ליצירה היא זכות היוצר -

- (1) כי שמו ייקרא על יצירתו בהיקף ובמידה הראויים בנסיבות העניין;
- (2) כי לא יוטל פגם ביצירתו ולא ייעשה בה סילוף או שינוי צורה אחר, וכן כי לא תיעשה פעולה פוגענית ביחס לאותה יצירה, והכל אם יש באילו מהם כדי לפגוע בכבודו או בשמו של היוצר."

כיוון שמאגרי נתונים לא מציינים ברגיל את שם היוצרים בתכנים הכלולים בהם, וכיוון שתהליך העיבוד של המאגר עשוי להכניס ביצירה שינויים משינויים שונים, עשוי המיזם להיות חשוף גם לטענה בדבר הפרת סעיף 46 לחוק. מדובר במצב מורכב מבחינה מעשית, מכיוון שהזכות המוסרית אינה ניתנת להעברה לצד שלישי, והיא נותרת בידיו של יוצר היצירה גם אם נעשתה

מדינת ישראל

משרד המשפטים

עסקה בזכות היוצרים.⁴¹ עבור מיזמי למידת מכונה, משמעות הדבר היא חשיפה משפטית הן מצד בעל זכות היוצרים ביחס לזכות הכלכלית שנדונה לעיל והן מצד היוצר ביחס לזכות המוסרית.

על אף האמור, סבורנו כי במקרה הרגיל, אין ממש בטענות בדבר הפרת הזכות המוסרית במיזמים של למידת מכונה. הזכות המוסרית אינה זכות מוחלטת.⁴² ראשית, אין היא חלה על כל סוגי היצירות.⁴³ שנית, כמצוטט לעיל, זכות היוצר כי שמו יקרא על יצירתו (זכות הייחוס) תלויה "בהיקף ובמידה הראויים בנסיבות העניין", וזכותו של היוצר כי יצירתו לא תשונה (זכות השלמות) תחול רק אם יש בשינוי "כדי לפגוע בכבודו או בשמו של היוצר".⁴⁴ אמנם, פסיקת בתי המשפט נוטה לתפיסה מרחיבה של הזכות המוסרית, בפרט באשר לזכות הייחוס.⁴⁵ ואולם, במקרה דנן, עמדתנו היא כי פרשנות התנאי בדבר ה"היקף והמידה הראויים" לצורך זכות הייחוס צריכה להיות צרה ביותר, וכי שינוי יצירה סטנדרטי בתהליך יצירת מאגר הנתונים, ככל שיהיה, צריך להתפרש ככזה שאין בו "כדי לפגוע בכבודו או בשמו של היוצר". לטעמנו, העדרו של קהל אנשי מאיין הן את האינטרס של היוצר במתן קרדיט והן את החשש לפגיעה בכבודו עקב שינוי היצירה.⁴⁶

מן המקובץ עולה כי מיזמי בינה מלאכותית עלולים לעמוד בפני טענות הן בדבר הפרת זכות יוצרים של בעל זכות היוצרים והן בדבר הפרת הזכות המוסרית של היוצר ביחס לכל יצירה ויצירה הכלולה במאגר. זאת ועוד. הסעדים בגין הפרת זכויות יוצרים הם מרתיעים במיוחד, וכוללים בין היתר צווי מניעה ופיצויים ללא הוכחת נזק בגובה של עד מאה אלף שקלים בגין כל הפרה.⁴⁷ החשיפה המשפטית של מיזמי בינה מלאכותית בדיני זכויות יוצרים היא אפוא גבוהה, אלא אם עומדים הם בתנאי ההסדרים של שימושים מותרים ביצירות. לדיון בהסדרים אלה נפנה כעת.

2. הסדרי השימושים המותרים

לצד סעיף 11 שנדון לעיל, נקבעו בחוק סייגים לזכויות הבלעדיות של בעלי זכות יוצרים. הסדרים אלה, המנויים בסעיפים 19-32 לחוק, מתירים שימוש ביצירות מוגנות ללא הרשאה מאת בעל זכות היוצרים וללא מתן תמורה בגין השימוש.⁴⁸ כפי שנראה להלן, על אימון מערכות של למידת מכונה עשויים לחול שלושה מסייגים אלה: שימוש הוגן (סעיף 19), שימוש אגבי ביצירה (סעיף 22) והעתקה זמנית (סעיף 26). כפי שידון להלן, המסגרת המתאימה ביותר לדיון היא הסדר השימוש ההוגן, אך לעתים עשוי להיות רלוונטי אחד או יותר מן ההסדרים האחרים.

⁴¹ ראו ס' 45 לחוק זכות יוצרים.
⁴² ראו גרינמן, לעיל ה"ש 12, בעמ' 823-884; יהושע ויסמן "הזכות האישית (droit moral) בדיני זכויות יוצרים" מחקר משפט 51 (1989); ג' טדסקי "הקניין הרוחני וזכויות אישיות" משפטים י 392 (התש"מ); פסח, לעיל ה"ש 10 (י"א)יש לתחום ולגדור את היקף הפריסה של זכות איהסילוף"; קים טרייגר בר-עם "הזכות המוסרית, הגנת הסבירות והאיזונים החיוניים" עלי משפט ח 237, 238 (2010) ("הזכות המוסרית לשלמות היצירה אינה זכות מוחלטת"). אין הרמוניזציה בינלאומית בנושא של זכויות מוסריות. ראו לדוגמה סעיף 19(1) להסכם הטריפס (Related-Trade on Aspects of Intellectual Property Rights Agreement (1994)).

⁴³ ס' 45(א), 50(ב) לחוק זכות יוצרים: "ליוצר של יצירה אמנותית, יצירה דרמטית, יצירה מוסיקלית או יצירה ספרותית, למעט תוכנת מחשב, שיש בה זכות יוצרים, תהיה ביחס ליצירתו זכות מוסרית, למשך תקופת זכות היוצרים באותה יצירה; ואולם, לעניין יצירה שהיא גופן לא תחול הזכות כאמור בסעיף 46(1)."

⁴⁴ ראו ס' 46 לחוק זכות יוצרים. ראו גם Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, art. 5(3)(a), 2001 O.J. (L 167), 10 (EC) ("turns out to be impossible") ("להלן: "הדירקטיבה Infosoc"). הדירקטיבה יושמה באופן מגוון במדינות האיחוד. החוק האיטלקי למשל מחייב מתן קרדיט ליוצר גם בהעתקה לצרכי מחקר. ראו (1) Legge diritto d'autore, Art. 70.

⁴⁵ ראו עניין קימרון, לעיל ה"ש 26; ע"א 782/87 אלחנני נ' עיריית תל-אביב-יפו, פ"ד מו(3) 529, 537-539 (1992); ע"א 3422/03 Krone AG נ' ענבר פלסטיק משוריין, פ"ד נט(4) 365, 375-376 (2005); ת"א (מחוזי ת"א) 1299/04 קוק נ' סיון אין האוס בע"מ (נבו 13.2.2008); ת"א (מחוזי-ים) 8303/06 מחולה המרכז למחול בע"מ נ' כהן, פס' 22-23 (פורסם בנבו, 14.8.2008); ת"א (מחוזי-חי') 10634-01-12 עיראקי נ' ח'טיב, פס' 5 לפסק הדין (נבו 11.2.2015); טוני גרינמן, "הזכות המוסרית - מ-Droit Moral ל-Moral Rights", יוצרים זכויות - קריאות בחוק זכות יוצרים 458 (מיכאל בירנהק וגיא פסח עורכים, 2008) ("אין בפסיקה דוגמות רבות למקרים שבהם נמצא כי לא הייתה חובה לאזכר את שם היוצר בנסיבות המקרה"). יובהר כי אין באמור משום הבעת הסכמה לפרשנות מרחיבה של הזכות המוסרית. דיון בסוגיה זו חורג ממסגרת חוות הדעת.

⁴⁶ ראו גם European Parliament, The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects - In Depth Analysis, 2018, [https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA\(2018\)604941_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA(2018)604941_EN.pdf), 11 (מציין בקשר לזכות השלמות כי- "impossible to fulfill this requirement").

⁴⁷ ראו ס' 56 לחוק זכות יוצרים.
⁴⁸ ראו ס' 19-32 לחוק זכות יוצרים.

מדינת ישראל

משרד המשפטים

א. שימוש הוגן

דוקטרינת השימוש ההוגן מהווה את מנגנון האיזונים המרכזי בחוק בין אינטרס ההגנה על יצירות לבין האינטרס הציבורי בהגברת הגישה ליצירות.⁴⁹ בישראל מעוגנת דוקטרינת השימוש ההוגן בסעיף 19:

19. (א) שימוש הוגן ביצירה מותר למטרות כגון אלה: לימוד עצמי, מחקר, ביקורת, סקירה, דיווח עיתונאי, הבאת מובאות, או הוראה ובחינה על ידי מוסד חינוך.
- (ב) לצורך בחינה של הוגנות השימוש ביצירה לעניין סעיף זה, יישקלו, בין השאר, כל אלה:
- (1) מטרת השימוש ואופיו;
 - (2) אופי היצירה שבה נעשה השימוש;
 - (3) היקף השימוש, מבחינה איכותית וכמותית, ביחס ליצירה בשלמותה;
 - (4) השפעת השימוש על ערכה של היצירה ועל השוק הפוטנציאלי שלה.
- (ג) השר רשאי לקבוע תנאים שבהתקיימם ייחשב שימוש לשימוש הוגן.

כפי שעולה מסעיף 19, כדי ששימוש ביצירה מוגנת יחסה תחת הגנת השימוש ההוגן, עליו לעמוד בשני תנאים מצטברים. ראשית, עליו להיעשות למטרה מאלה המנויות בסעיף קטן א (או למטרה "כגון אלה").⁵⁰ שנית, עליו להבחין על פי מבחני השימוש ההוגן המנויים בסעיף קטן ב.⁵¹

נעבור אפוא לבחינת סעיף 19(א). נראה כי יצירת מאגר נתונים לצורך אימון מכונה מתאימה היא הן להגדרת "לימוד עצמי" והן להגדרת "מחקר" בסעיף 19(א), ולכל הפחות נעשית למטרות "כגון אלה".⁵² ראשית, למידת מכונה נופלת לגדר "לימוד עצמי", בהיותה לימוד של המכונה את עצמה. אכן, לימוד מכונה מקביל הוא לתהליך אנושי של לימוד עצמי אינדוקטיבי. זו גם זו נעשות באמצעות למידה מדוגמאות. ההבחנה בין למידה אנושית ללמידת מכונה אינה אלא ההליך הטכני-טכנולוגי של אופן הלמידה: בעוד המוח האנושי יכול ללמוד מדוגמאות תוך התקלות אקראית בהן מבלי שיהיה צורך לייצר עבורו מאגר נתונים ייעודי, יכול המחשב "ללמוד" אך ורק ממאגר נתונים המאורגן בצורה מסויימת. כיוון שהבחנה זו עוסקת בתהליך הלמידה עצמו ואינה רלוונטית לקיום תכלית הסעיף, ראוי לפרש את המושג "לימוד עצמי" ככולל גם למידת מכונה. פרשנות זו תואמת

⁴⁹ ראו עניין **Premier League**, לעיל ה"ש 4, פס' 17-18 לפסק דינו של המשנה לנשיאה ריבלין: "מדובר [...] בכלי שנועד לאפשר - במידת הצורך - לבצע איזון בין הרצון להעניק תמריץ ליוצרים להפקת יצירות חדשות, לבין השאיפה להעשיר את מגוון היצירות במרחב הציבורי[...]"; אורית פישמן-אפורי **היצירה הנגזרת בדיני זכויות יוצרים** 332 (2005) ("בהיות הגנת השימוש ההוגן כלי לאיזון בין אינטרס בעל הזכויות לבין אינטרס הציבור, מטבע הדברים קיימת מחלוקת בשאלה אם מושג באמצעותה האיזון הרצוי, ואם די בו לשם שמירה על אינטרס הציבור."); ניבה אלקין-קורן, "זכויות משתמש" **יוצרים זכויות - קריאות בחוק זכות יוצרים** 327 (מיכאל בירנהק וגיא פסח עורכים, 2008); Michael Birnhack, *Judicial Snapshots and Fair Use Theory*, 5 *QUEEN MARRY J. INTELL. PROP.* 264, 264 (2015).

⁵⁰ פרק השימושים המותרים החליף רשימה סגורה של שימושים מותרים שנהגה בחוק הקודם - ס' 12 לחוק זכות יוצרים, 1911, חא"י ג 2633(ט). שינוי זה נתפס כביטוי לכוונת המחוקק להרחיב שימוש חופשי ביצירות. ראו גרינמן, לעיל ה"ש 12, בעמ' 327 ("ההוראות לעניין זה [השימושים המותרים] קבועות בעיקרן בפרק ד' לחוק, שכותרתו "שימושים מותרים". הוראות אלה הינן החידוש המהותי ביותר שהונהג בחוק זכות יוצרים בהשוואה לדין הקודם והן גרמו, בעקבות חקיקתו, לשינוי באיזון בין זכויות היוצרים לבין האינטרסים של המשתמשים ביצירה."); יהושע ויסמן "קריאה השוואתית: מאפיינים של חוק זכות יוצרים, התשס"ח-2007" **יוצרים זכויות - קריאות בחוק זכות יוצרים** (תשס"ט) 61, 80 ("... הרשימה הפתוחה של הגנת השימוש ההוגן, כמו ההוראה השוללת את הגנת זכות יוצרים מהרעיונות שביצירות, מצמצמת באופן ניכר את זכויותיהם של היוצרים, תוך הדגשה יתרה של האינטרס הציבורי בשימוש חופשי ביצירות התרבות לעומת האינטרסים של היוצרים בפרי יצירתם."). מהלך דומה של מעבר מרשימה סגורה לפתוחה נעשה גם באוסטרליה. ראו *Emily Hudson, Implementing Fair Use in Copyright Law: Lessons from Australia*, 25(3) *I.P.J.* 201 (2013).

⁵¹ בפועל, המשוכה של סעיף 19(א) אינה משמעותית בפסיקה (ולעתים אף לא נבחנת כלל), ועיקר הבחינה מתמקד בבחינת סעיף 19(ב). ראו למשל עניין **Premier League**, לעיל ה"ש 4, פסקאות 20-19 לפסק דינו של השופט ריבלין ("ימדלג" על בחינת סעיף 10(א); ע"א 7996/11 **סייפקום בע"מ נ' עופר רביב**, פס' 37 לפסק דינו של השופט דנציגר (נב) 18.11.2013. (להלן: "עניין סייפקום") (כנ"ל). ראו גם גרינמן, זכויות יוצרים, 406-407 (2008) ("... בדומה לבתי המשפט האמריקניים - יש לבחון את השאלה בשלב אחד ורציף, שבא לבחון את הוגנות השימוש, כאשר מטרת השימוש היא אחד השיקולים שיובאו בחשבון במסגרת המבחן הראשון, שעניינו מטרת השימוש ואופיו."). אך ראו ע"א 3425/17 **Societe des Produits Nestle נ' אספרסו קלאב בע"מ** (7.8.2019) (להלן: "עניין אספרסו קלאב"), פסקה 21 לפסק דינו של השופט הנדל ("לשון סעיף 19(א) אינה מאפשרת לוותר על מבחן המטרה כמבחן סף עצמאי").

⁵² ראו ס' 19(א) לחוק זכות יוצרים.

מדינת ישראל

משרד המשפטים

את המציאות הטכנולוגית העכשווית ועולה לדעתנו בקנה אחד עם הפסיקה בישראל, שפירשה את המושג "לימוד עצמי" בהרחבה.⁵³

מלבד לימוד עצמי, נופלת יצירה של מאגר נתונים לדעתנו גם לגדר "מחקר", הן בשל התהליך של ניתוח, בחירה ואפיון של התכנים הנדרשים, מיונם ו"תיגום", והן בשל כך שהאופן בו מחשבים לומדים עודו מהווה מושא למחקר בעצמו.⁵⁴ עוד יצויין כי הן המושג "מחקר" והן המושג "לימוד עצמי" פורשו בהרחבה ניכרת וכוללים גם שימושים מסחריים.⁵⁵ מכל מקום, מתיר הסעיף שימוש בחומרים מוגנים גם למטרות "כגון אלה". לכן, גם לולא עמדה יצירתו של מאגר הנתונים בבחינה מילולית-קפדנית של הגדרת "לימוד עצמי" או "מחקר" (על אף פרשנותם הרחבה בפסיקתנו), הרי קרובה היא מספיק אליהן על מנת לעמוד, לדעתנו, ללא ספק של ממש, בתנאי סעיף 19(א).

משנעברה המשוכה של סעיף 19(א), יש לבחון אם מתקיימים תנאיו של סעיף 19(ב). סעיף זה מונה ארבעה שיקולים שיש לבחון, "בין היתר", "לצורך בחינה של הוגנות השימוש ביצירה".⁵⁶ הגם שניתן לבחון שיקולים נוספים, ארבעת השיקולים המנויים בחוק מהווים את עמודי התווך בבחינת הוגנות השימוש. להלן יוצע ניתוח של ארבעת שיקולי השימוש ההוגן בהקשר הקונקרטי של למידת מכונה לצד שיקולים. ניתוח זה יביא למסקנה כי במקרה התדיר של למידת מכונה, תחסה בניית מאגר הנתונים תחת הסדר השימוש ההוגן.

(1) מטרת השימוש ואופיו

השיקול הראשון המנוי בסעיף 19(ב) עניינו "מטרת השימוש ואופיו".⁵⁷ מטרת השימוש בענייננו היא יצירה של מאגר נתונים אפקטיבי עבור למידת מכונה. יצירת מאגר כאמור היא קריטית, כאמור, כדי שמערכות בינה מלאכותית יבנו בצורה אפקטיבית, בטוחה ותחרותית. אופי השימוש נבחן בפסיקה בעיקר לאור השאלה אם השימוש שנעשה ביצירה הוא שימוש ביצירה כפי שהיא, או שהוא מהווה שימוש טרנספורמטיבי, קרי שימוש המשנה את היצירה, את משמעותה או את הקשרה. ככל שהשימוש הוא טרנספורמטיבי יותר, כך נחשב הוא "הוגן" יותר, על בסיס ההנחה ששימוש טרנספורמטיבי מעשיר את עולם היצירה, והציבור יוצא נשכר.⁵⁸

שאלת הטרנספורמטיביות בהקשר של למידת מכונה היא מורכבת במיוחד. טכנית, למאגר הנתונים הראשוני מועתקות יצירות מבלי לערוך בהן שינוי משמעותי.⁵⁹ ברם, בחינה הוליסטית של תהליך היצירה של מאגרי נתונים מלמדת שימוש כזה הוא טרנספורמטיבי ככל שניתן להעלות על הדעת, לאור עוצמת השינוי הגלום בו מבחינת הקשרו. מערכת נהיגה אוטונומית, למשל, "צופה" בסרטים כדי שהרכב ילמד את עקרון "קביעות האובייקט" וידע לצפות שהולך רגל שנעלם מאחורי

⁵³ ראו למשל ת"א (מחוזי-ים) 8303/06 מחולה המרכז למחול בע"מ נ' כהן תק-מח 37396 (3)08 (2008), לפיו הוראת יצירת מחול תוך ביצועה בפני ועל ידי תלמידים מהווה לימוד עצמי.

⁵⁴ ראו גם Dan Brown, Lauren Byl & Maura R. Grossman, *Are machine learning corpora "fair dealing" under Canadian law?* 159 (2021) (להלן: "להלן": *fair dealing*) (מציין כי גם על פי חריגי זכויות היוצרים בדן הקנדי, יש לפרש את המושג מחקר בהרחבה ומצטט את פס"ד, CCH Cdn. Ltd. v. LSUC (2004), 266 F.T.R. 159 (FC) לפיו "Research" must be given a large and liberal interpretation in order to ensure "that users' rights are not unduly constrained."

⁵⁵ ראו למשל גרינמן, לעיל ה"ש 12, בעמ' 414 ("השאלה שיש לשאול... [היא] אם חקר היוצר את החומר, האיר, העיר או ביקר אותו, אם הם התייחס באופן אחר העשוי להיקרא "מחקר", לחומר עצמו."); ראו גם ת"א (שלום ת"א) 24595/97 בס נ' כתר הוצאה לאור בע"מ, פ"מ תשס"ב (3) 337, פס" 61 לפסק דינו של השופט שחר (2002) (קובע ששימוש ביצירות נעשה כחלק מעבודת מחקר כיוון שהיצירות האירו את התוכן המילולי); ת"א (שלום י-ם) 8397/98 ביטון נ' סולטון, תק-של (2)00 9883 (2000), אושר ברע"א 1108/02 ביטון נ' סולטון, תק-על (2)02 1110 (2002) (מכיר במטרת מחקר גם בהקשר מסחרי). גם באיחוד האירופי חריג המחקר, הקבוע בס' 3(5) לדירקטיבה, פורש כמסגרת נורמטיבית ראויה לסוגיה של TDM, החופפת במידה רבה לסוגיה זו. ראו להלן פרק ד.

⁵⁶ ראו ס' 19(ב) לחוק זכות יוצרים.
⁵⁷ ראו ס' 19(ב) לחוק זכות יוצרים.

⁵⁸ ראו למשל עניין Premier League, לעיל ה"ש 4, פסקה 20 לפסק דינו של המשנה לנשיאה ריבלין ("התפיסה היא שקל יותר להכיר בשימוש טרנספורמטיבי כ'הוגן', שכן הוא מגשים את תכליתו של ההיתר – לעודד את היצירה ולהעשיר את מאגר הידע המצטבר בחברה."); עניין *אספרסו קלאב*, לעיל ה"ש 51, פסקה 28 לפסק דינו של השופט הנדל ("שיקול הטרנספורמטיביות נתפס כיום כנתון המרכזי ביותר בפסיקה האמריקאית, וגם בפסיקה בישראל הוא הוכר כשיקול מרכזי בגדרי בחינת אופי השימוש"); עניין *סייפוקס*, לעיל ה"ש 51, פסקה 37 לפסק דינו של השופט דנציגר; רע"א 7774/09 ויינברג נ' ויסהוף (נבו 28.8.2012), פסקה 22 (להלן: "עניין ויינברג"). ראו גם Suntrust Bank v. Houghton Mifflin Co., 268 F.3d 1257 (2001) (מנתח את הרומן שכתבה אליס רנדל *The Wind Done Gone*, המתאר את הסיפור "חלף עם הרוח" מנקודת המבט של השפחה השחורה כפארודיה על *Gone with the Wind* של מרגרט מיטשל, ורואה בשימוש ביצירה המקורית לצורך הרומן משום שימוש טרנספורמטיבי).
⁵⁹ ראו לעיל פרק ב.

מדינת ישראל

משרד המשפטים

עץ ישוב ויצוף, ולא על מנת להתרשם מהאיכות האסתטית של הקומפוזיציה בסרט. באופן דומה, כאשר מערכת NLP "קוראת" טקסט, אין היא חותרת להעמיק בתזה הכתובה בו אלא, לדוגמה, ללמוד לזהות יישויות במשפט.⁶⁰ נכון הדבר, הקומפוזיציה בסרטים והתאורה שבה צולמו, כמו גם גיוון המשפטים ובחירת המילים ביצירות המקוריות, מהווים כלי למידה חשובים עבור המערכת. ועם זאת, השימוש שעושה בהם המערכת לבסוף שונה הוא בתכלית השינוי מהשימוש המקורי.⁶¹

ודוק. במערכות מסוימות של למידת מכונה, השימוש ביצירות אינו טרנספורמטיבי כבאחרות. כך, למשל, ניתן ללמד מערכות להפיק תוצר שנושא דמיון ניכר ליצירות שנכללו במאגר הנתונים.⁶² הדבר יכול להתרחש בכוונת מכוון (למשל, כאשר מערכת מתוכננת לחקות יוצר או ז'אנר מסוימים)⁶³ או בלעדית, כאשר מאגר המידע אינו מגוון דיו.⁶⁴ במקרים כאלה, יתכן שלא יהיה מנוס מלקבוע כי השימוש שנעשה ביצירות אינו טרנספורמטיבי.

לבסוף, בחינת אופי השימוש כוללת גם את השאלה אם השימוש הוא מסחרי. היותו של שימוש מסחרי מהווה משקל נגד בהכרעה אם מדובר בשימוש הוגן,⁶⁵ אך אין בו כדי לשלול את תחולת הדוקטרינה מינייה וביה, בעיקר כשמדובר בשימוש טרנספורמטיבי.⁶⁶

לאור הניתוח לעיל, עמדתנו היא שהמקרה הפרדיגמטי של למידת מכונה עומד בקריטריון הקבוע בסעיף 19(ב)(1), בהיותו בעל ערך מבחינה חברתית וטרנספורמטיבי – המשנה בתכלית את הקשר של היצירה המקורית. עם זאת, במצבים בהם המיזם אינו טרנספורמטיבי, ובעיקר כאשר הוא גם מסחרי, טיטה הכף לטובת הכרעה שסעיף 19(ב)(1) מצביע לכיוון אי-הוגנות של השימוש.

(2) אופי היצירה שבה נעשה השימוש

השיקול השני בבחינת הוגנות השימוש בוחן את אופי היצירה שבה נעשה השימוש: באיזו יצירה מדובר ובאיזו מידה מצויה היא בליבת דיני זכויות היוצרים. כך למשל, שימוש ביצירה ספרותית או מוסיקלית עשוי להבחן ביתר קפדנות לעומת שימוש בשידור או בהקלטה. כמו כן, שימוש ביצירות המבוססות על עובדות ונתונים או ביצירות הנושאות אופי מחקרי יסווגו כשימוש הוגן ביתר קלות לעומת יצירות שהן פרי דמיונו של היוצר. במידה רבה, שיקול זה נחשב כמשני בחשיבותו בין מבחני השימוש ההוגן,⁶⁷ ומכל מקום, קיים קושי אינהרנטי להעריך את התקיימותו באופן קטגורי בשל המגוון האינסופי של שימושי בינה מלאכותית ושל מאגרי הנתונים שיוכנו עבורם.

⁶⁰ ראו גם Mark A. Lemley & Bryan Casey, *Fair Learning*, 99(4) TEX. L. REV. 743 (2021).
⁶¹ ראו גם Kelly v. Arriba Soft Corp., 336 F.3d 811 (9th Cir. 2003); Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1176 (9th Cir. 2007). הקובעות כי הכללת תמונות מלאות - אך מוקטנות (thumbnails) - במנועי חיפוש נחשבת לשימוש טרנספורמטיבי בשל הפונקציונליות השונה של השימוש.

⁶² כך למשל נטען בתובענה ייצוגית שהוגשה לאחרונה בארה"ב בעניין GitHub Copilot - כלי לכתיבת קוד מבוסס בינה מלאכותית שפיתחה מייקרוסופט - בין היתר בטענה שהפלט של המכונה יכול להפיק קטעי תוכנה זהים לקטעי התוכנה שנכללו במאגר הנתונים. ראו Complaint, Class action, J. DOE et al. v. Github, Inc. et al., Case N.D. Cal. (filed 11.3.22). זמין לצפייה ב- https://githubcopilotlitigation.com/pdf/1-0-github_complaint.pdf.

⁶³ ראו לדוגמה שם, וכן להלן ה"ש 136-137.

⁶⁴ השימוש במאגרי מידע לא מגוונים דיים עשוי לנבוע מסיבות שונות, כגון קיומם של מאגרים חלקיים וגישה מוגבלת למאגרים קיימים, וליצר מגוון בעיות מעבר לענייני זכויות יוצרים, החל בפגיעה באפקטיביות המערכת ובורסטיטיליות שלה וכלה בהטייה לרעת קבוצות שאינן מיוצגות במאגר. ראו למשל Amanda Levendowski, *How Copyright Law* Kirsten Lloyd, *Bias ; Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579 (2018). Available at: *Amplification in Artificial Intelligence Systems*, 2 (2018). Available at: Bert Heinrichs, *Discrimination in the Age of Artificial Intelligence*, 37 AI & SOC'Y 143, 150-151 (2022). <https://doi.org/10.48550/arXiv.1809.07842>.

⁶⁵ ראו למשל רע"א 92/2687 גבע נ' חברת וולט דיסני, פ"ד מח(1)251, 271 (1993), בעמ' 276; עניין Premier League לעיל ה"ש 4, פסקה 20 לפסק דינו של השופט ריבלין.

⁶⁶ ראו למשל עניין Premier League, שם (יעצם קיומו של שימוש מסחרי אינו שולל את תחולתה של הגנת השימוש ההוגן. אולם מקובל לומר כי ככול שאכן מדובר בשימוש מסחרי, נחלש הטיעון בדבר השימוש ההוגן); עניין אספרסו קלאב, לעיל ה"ש 51, פסקאות 30-31 ("שליטת ההגנה משימוש מסחרי, בוודאי למטרות רווח, עלולה לרוקן מתוכן חלקים נרחבים מהסדר השימוש ההוגן, ולפגוע בתכליות החברתיות החשובות שלשמו נועד"). ראו גם Campbell v. Acuff-Rose Music Inc., 114 S.Ct. 1164, 1176 (1994); Authors Guild v. Google, Inc., 804 F.3d 202, 219 (2d Cir. 2015) ("[there is] no reason . . . why [a defendant's] overall profit motivation should prevail as a reason for denying fair use over its highly convincing transformative purpose, together with the absence of significant substitutive competition, as reasons for granting fair use"). (להלן: "פרשת Google Books").

⁶⁷ ראו למשל עניין Premier League, שם, בו שיקול זה אף לא נבחן. ראו גם עניין אספרסו קלאב, לעיל ה"ש 51, פסקה 34 לפסק דינו של השופט הנדל ("השיקול של אופי היצירה המוגנת יתפוס מקום רב רק במקרים בעלי פן

מדינת ישראל משרד המשפטים

(3) היקף השימוש ביצירה מבחינה איכותית וכמותית

השיקול השלישי לבחינת הוגנות השימוש עניינו היקף השימוש ביצירה מבחינה איכותית וכמותית.⁶⁸ הכוונה היא לבחון מה - וכמה - ניטל מהיצירה המקורית במסגרת השימוש בה. בהקשר של למידת מכונה, לרוב מועתקות יצירות בכללותן למאגרי המידע בשלב של יצירת המאגר, ולכאורה מהווה הדבר אינדיקציה לשימוש שאינו הוגן. עם זאת, לעתים קרובות (אם כי לא תמיד), יצירות מועתקות למאגר הנתונים לא על מנת לעשות שימוש בביטוי היצירתי הגלום בהן אלא כדי לאפשר גישה לאלמנטים בלתי מוגנים הכלולים ביצירה, כגון, עובדות, רעיונות, מבנה תחבירי וכיוצא באלה.⁶⁹ העתקת היצירה נעשית רק מפני שמחשב, במובחן מן אנוש, אינו יכול "לקרוא" חומרים שאינם מוגנים מבלי לבצע קודם כל העתקה של היצירה בכללותה לתוך מאגר נתונים קריא עבורו.⁷⁰ מדובר במגבלה טכנית הנובעת מארכיטקטורה הכרחית של מערכות לומדות, ופרשנות תכליתית שמבקשת לאפשר למידת מכונה חייבת להביא אותה בחשבון. לאור זאת, יש לנתח את היקף השימוש ביצירה לא רק לאור העתקת היצירה בשלמותה אלא לאור זאת שהלמידה עצמה צפויה להתבצע בסופו של דבר רק מחלק מכל יצירה ולאור בחינת האלמנטים ביצירה מהם מתבצעת הלמידה. כאשר הלמידה נעשית בעיקר מהאלמנטים ביצירות שאינם מוגנים, צריך שיקול זה לדעתנו לחזק את טענת השימוש ההוגן.

(4) השפעת השימוש על השוק של היצירה ועל השוק הפוטנציאלי שלה

השיקול הרביעי לבחינת שימוש הוגן עוסק בהשפעת השימוש על השוק של היצירה. כאן, לדעתנו, נוטה הכף לטובת הוגנות השימוש בהקשר של למידת מכונה. ראשית, במצב השוק דהיום, הרחבת הגישה של מכוונות ליצירות לא תפגע בשווקים קיימים של בעלי זכויות יוצרים שכן שווקים כגון אלה אינם בנמצא. אכן, גם לו ביקשו יוצרים של מערכות למידת מכונה לרכוש רשיונות עבור כל אחד מהתכנים המרכיבים את מאגר הנתונים, הדבר אינו בגדר האפשר מבחינה מעשית. אין בנמצא פלטפורמה המציעה רשיונות כאמור, ואילו היקף התכנים הנדרשים, פרישתם הגאוגרפית והעדר רישום של בעלויות וזכויות ביצירות מאיינים כל אפשרות לפעול להשגת רשיונות באופן פרטני.⁷¹ ברי כי גם בהקשר זה, מתקיים יחס ישר בין רמת הגיוון של המאגר לבין המסקנה בדבר העדרו של שוק לרשיון ביצירות. כאשר מדובר ביצירה של יוצר בודד, חזקה כי אין מניעה להתקשר עמו בהסכם פרטני לשימוש ביצירותיו לצורך הקמת מאגר מידע.

אכן מעבר לבחינת קיומו של שוק במובן הסטטי, דהיינו - הבחינה אם שוק כאמור ישנו בנמצא - יש לבחון במסגרת השיקול הרביעי גם השפעה על שוק היצירות במונחים דינאמיים, דהיינו, התכנות של הוצרות שוק רלוונטי בעתיד.⁷² הרי אין לחדד כי פרשנות לפיה למידת מכונה מהווה שימוש הוגן תקבע את השוק של בעלי הזכויות לשווקים הקיימים כיום, ותקשה על הוצרותו של שוק לשימוש החדש שנמצא ליצירות כבסיס ללמידת מכונה. עם זאת, לדעתנו, גם בחינה דינאמית מובילה למסקנה כי כלילת יצירות במאגר מידע לצורך למידת מכונה מהווה שימוש הוגן. הטעם לכך הוא שכשלי שוק מבניים מולידים סבירות נמוכה ביותר להוצרות שוק אפקטיבי לאספקת רשיונות בזירה זו. בפרט, בעיות מידע וקוארדינציה, הנובעות בין היתר מהעדר רישום של זכויות ביצירות, ועלויות עסקה מאסיביות, הנובעות בין היתר מהפיזור של תכנים (ושל הדין החל עליהם) על פני הגלובוס, צפויים למנוע הוצרותו של שוק רישוי יעיל לצורך למידת מכונה.

ייחודי). בחינה אמפירית של השימוש שעושים בתי המשפט במבחני השימוש ההוגן בארצות הברית מובילה למסקנה זהה. ראו למשל 156, *Barton Beebe, An Empirical Study of U.S. Copyright Fair Use Opinions, 1978-2005*, 156 LEWIS & CLARK L. REV. 715 (2008); Neil W. Netanel, *Making Sense of Fair Use*, 15 LEWIS & CLARK L. REV. 715 (2011); Mathew Sag, *Predicting Fair Use*, 73 OHIO ST. L.J. 47 (2012).

⁶⁸ ראו 539, 556 (1985) 471 U.S. *Harper & Row, Publishers, Inc. v. Nation Enters.*, (שם נפסק כי העתקה של 300 מילים בלבד מתוך ביוגרפיה של 400 עמודים נחשבה לנטילה של היקף משמעותי מבחינה איכותית).

⁶⁹ ראו ס' 5 לחוק זכות יוצרים. ראו גם Daniel Gervais, *The Machine as Author*, 105 IOWA L. REV. 24 (2019) ("TDM is looking, if anything, for ideas embedded in copyright works").

⁷⁰ ראו גם לעיל ה"ש 16-17 וטקסט מלווה.

⁷¹ ראו גם פרשת *Google Books*, לעיל, ה"ש 66, בה נפסק כי פרוייקט הספרים של גוגל לא הזיק לבעלי הזכויות, ולמעשה תרם להם באמצעות פרסום יצירותיהם.

⁷² בראשית המילניום בתי משפט בארצות הברית שקלו את הרעיון כי היתכנותו של שוק צריכה לשלול שימוש הוגן. ראו *Princeton Univ. Press v. Mich. Document Servs., Inc.*, 99 F.3d 1381, 1388 (6th Cir. 1996); *Am. Geophysical Union v. Texaco, Inc.*, 60 3d 913, 930-31 (2d Cir. 1994) ("[A]n unauthorized use should be considered 'less fair' when there is a ready market or means to pay for the use."). עמדה זו עוררה ביקורת עזה בספרות. *James Gibson, Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 YALE L.J. 882, 931-35 (2007); *Wendy J. Gordon, The 'Why' of Markets: Fair Use and Circularity*, POCKET PART L.J. 882, 931-35 (2007); <http://yalelawjournal.org/2007/4/25/gordon.html>, 371 (2007); ניבה אלקין-קורן "הסדרה עצמית של זכויות יוצרים בעידן המידע" עלי משפט ב 319, 341 (2002).

מדינת ישראל משרד המשפטים

הראיה, כשלים כגון דא מנעו עד היום התפתחות של שווקים יעילים בהקשרים רבים מספור של שימוש בזכויות יוצרים בעולם הדיגיטלי.⁷³

ודוק. הקביעה כי הסדר השימוש ההוגן חל בסוגיה דנן לא שולל מיניה וביה הווצרותו של שוק כאמור.⁷⁴ אדרבה, הגנת השימוש ההוגן עשויה דווקא לתמרץ היווצרותו של שוק. ככל שיווצר שוק יעיל לרישוי יצירות לצורך למידת מכונה, עשוי השיקול דנן להפוך את כיוונו ולהחליש את טענת השימוש ההוגן בהקשרים המסויימים בהם נוצר שוק כאמור. יחד עם זאת, הקביעה לטובת הוגנות השימוש מאפשרת את שגשוגה של תעשיית הבינה המלאכותית כל עוד שווקים כאמור אינם נבמצא ובשים לב לקשיים האינהרנטיים ביצירת שווקים כאלה.⁷⁵

יתר על כן, הכרעה נגד החלת הדוקטרינה של שימוש הוגן בהקשר דנן היתה עלולה לייצר השלכות תחרותיות שליליות על השוק בשני היבטים לפחות. ראשית, עלויות העסקה הגבוהות היו מייצרות סממי כניסה גבוהים לתחום הבינה המלאכותית עבור חברות הזנק צעירות. חברות עתירות משאבים יכולות לדלג על המשוכה של אי בהירות משפטית ביתר קלות. שני טעמים לדבר. ראשית, חברות טכנולוגיה המצויות בשוק מזה זמן רב מחזיקות במאגרי מידע משל עצמן. כך, למשל, גוגל ביססה מערכת NLP על למידת תכתובות האימייל של מליוני משתמשיה על פני עשרות שנים. למידע זה אין מקבילה בחברות הזנק צעירות, הנדרשות להשיג את המידע בדרך אחרת או להשתמש במידע סינתטי, דבר שעשוי לפגוע באיכות מוצריהם. שנית, חברות עתירות משאבים יכולות לבצע רכישה של זכויות שימוש במידע גם מעל הצורך. בדרך זו, אין הן **פותרות** את אי הבהירות אך הן **נפטרות** ממנה. כבר היום, רכישת רשיונות לביצוע פעולות שניתן לעשותן גם ללא רשיון מכוח הסדר השימוש ההוגן מהווה פרקטיקה מקובלת בקרב משתמשים עתירי משאבים ושונאי סיכון.⁷⁶ מסתמא, חברות הזנק לא יכולות להרשות לעצמן רכישות עודפות ממין זה. יתר על כן, התנהגות מגננתית כזו של בעלי זכויות עשויה לייצר החצנות שליליות שישפיעו לרעה על יכולתן של חברות צעירות לבצע שימושים אלה באופן חופשי. זאת הן דה-פקטו, מכיוון שחברות תוכן יפתחו ציפיה לקבל תשלום בגין שימוש כאמור והן דה-יורה, מכיוון ששאלת קיומו של שוק משפיעה על ההכרעה אם שימוש ביצירה מהווה שימוש הוגן.⁷⁷ מיזמי בינה מלאכותית עלולים היו "לנדוד" אפוא לתחומן של חברות גדולות, עם כל המחירים הנלווים של הדבר לעולם היזמות, בעולם בכלל ובישראל בפרט.⁷⁸

היבטים מעשיים של איסוף חומר מקוון מולידים בעיית ריכוזיות בהיבט נוסף. לו הכרענו כי איסוף תכנים ושימוש בהם לצורך למידת מכונה מחייבים רישוי, עשוי היה הדבר להגביר ריכוזיות בשוק התוכן, כיוון שהכרעה כאמור היתה מחזקת את מעמדם השוקי של גופי תוכן ריכוזיים על חשבון יוצרים אינדיווידואלים או לא מאוגדים.⁷⁹ אכן, לנוכח מגבלה משפטית על האפשרות לאסוף חומר מבוזר באופן מקוון, נוצר היה תמריץ חזק בשוק לאסוף חומרים ללמידת מכונה מגופי תוכן מאסיביים, כגון מוציאים לאור גדולים, תחנות רדיו, ארגוני ניהול משותף של זכויות וכיוצא באלה. תמריץ שוקי כזה עלול היה לפגום גם ברמת הגיוון של מאגרי נתונים.

לעומת מחירים אלה, הנזק לבעלי זכויות היוצרים כתוצאה מהכללת יצירותיהם במאגרי נתונים הוא מזער, אם בכלל קיים. ראשית, כאמור, אין נבמצא שוק רישוי יעיל לשימוש ביצירות לצורך למידת מכונה, ולכן גם לו הוכרע כי שימוש כאמור אינו מהווה שימוש הוגן, הרי שמלבד - אולי - פניה לגופי תוכן גדולים או ארגוני ניהול משותף, שברשותם תכנים מסוגים מסויימים בלבד, **כלל לא היה בידם של מיזמי למידת מכונה לרכוש רשיון שימוש לפעילותם**. יתר על כן, גם לו היה נבמצא שוק יעיל לרישוי יצירות עבור למידת מכונה, הרי מחיר השוק של הרשיון היה משקף את העובדה שערכה של כל יצירה אינדיווידואלית במאגר הנתונים הוא שולי ביותר. באיזון בין הרווח

⁷³ ראו למשל 36(2) COLUM. J. L. & ARTS 157, *Lital Helman, Fair Trade Copyright*.

⁷⁴ ראו למשל להלן ה"ש 76 וטקסט מלווה.

⁷⁵ ישנה בעגליות מובנית בשיקול הרביעי לבחינת הוגנות השימוש. שהרי, ככל שלא קיים שוק ליצירות, כך קטן הסיכוי לראות בשימוש בהן שימוש הוגן, אולם ככל ששימוש מסויים יוכר כהוגן, כך קטנה (אולי) הסבירות שיווצר עבורו שוק. לפיכך, לעתים קשה להימלט מהצורך בהכרעה נורמטיבית בשאלה האם ראוי ששוק כאמור ייווצר. למען הסר ספק יובהר כי חוות דעת זו נמנעת במכוון מהכרעה נורמטיבית כאמור.

⁷⁶ ראו James Gibson, *Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 YALE L.J. 882, 931-35 (2007); Wendy J. Gordon, *The 'Why' of Markets: Fair Use and Circularity*, POCKET PART 371 (2007), <http://yalelawjournal.org/2007/4/25/gordon.html>; Wendy J. Gordon & Daniel Bahls, *The Public's Right to Fair Use: Amending Section 107 to Avoid the "Fared Use" Fallacy*, 2007 UTAH L. REV. 619 (2007); אלקין-קורן, לעיל ה"ש 73.

⁷⁷ ראו סעיף 19(ב)4 לחוק זכות יוצרים.

⁷⁸ ראו JOSEPH SCHUMPETER, *BUSINESS CYCLES* 93 (1939) (לפיו חברות צעירות יזמיות מחברות מבוססות).
⁷⁹ ריכוזיות בשוקי תוכן מהווה תופעה ידועה בזירת זכויות היוצרים. ראו למשל ליאור ברוך, מעיין פרל וניבה אלקין-קורן, "תחרות ופער יזמות בשוק המוזיקה הדיגיטלי: חישוב מסלול מחדש", *מחקרי רגולציה ז* (2021), בעמ' 331.

מדינת ישראל

משרד המשפטים

הנמוך - אם בכלל - שצפויים היו בעלי זכויות יוצרים להפיק כתוצאה מפרשנות צרה לדוקטרינת השימוש ההוגן בהקשר דן לבין הרווח הכלכלי והתחרותי המשמעותי שיווצר כתוצאה מהחלה רחבה שלה, ברי בעינינו שיד האחרונה על העליונה.

נמצאנו למדים כי במרבית המקרים, בחינה מצטברת של שיקולי השימוש ההוגן תביא למסקנה כי שימוש בתכנים מוגנים לצורך למידת מכונה ייחשב לשימוש הוגן. כך, **מטרת השימוש ואופיו** מצביעים על שימוש טרנספורמטיבי למטרה רצויה, גם אם לעתים קרובות מסחרית; **אופי היצירה** עשוי להשתנות ממקרה למקרה ולא ניתן לקבוע בו מסמרות; **היקף השימוש** יצביע לכיוון הוגנות השימוש ברוב המקרים, ובפרט כאשר גם כשהיצירה מועתקת בשלמותה, המערכת לומדת ממאפייניה שאינם מוגנים; **והשפעה על השוק של היצירה** היא זניחה לכל היותר, הן לפי מצב הדברים בעת הזו והן לאור ניתוח מבני של שווקי התוכן בזירה המקוונת.

מעבר לארבעת הקריטריונים הללו, מקובל בפסיקתנו לבחון גם את השאלה אם ניתן קרדיט ליוצר כחלק משאלת השימוש ההוגן (פרשנות שאין בחוות דעת זו משום תמיכה בה).⁸⁰ יובהר כי על פי הניתוח דכאן, אין טעם בייחוס היצירה ליוצר בפני מכונה שעושה שימוש ביצירותיו.⁸¹ אשר על כן, אין בעובדה כי לא ניתן קרדיט ליוצר בתהליך יצירת המאגר כדי לשנות מהניתוח האמור בדבר היותה של למידת מכונה בגדר שימוש הוגן.⁸²

לעומת זאת, בהקשר של מאגרי מידע ללמידת מכונה, נכון להביא בחשבון שיקול מרכזי נוסף שאין לו מקום מובחן במסגרת ארבעת השיקולים המנויים בסעיף, והוא רמת הגיוון של מאגר המידע.⁸³ אכן, שונה מאגר המורכב מתכנים של יוצרים מוגוונים ממאגר המורכב מיצירתו של יוצר בודד שבאמצעותו לומדת המערכת את סגנונו המסויים של היוצר. בין רמת הגיוון של המאגר לבין הוגנות השימוש צריך שיתקיים יחס ישר, כך שכלל שמאגר המידע יהא מגוון יותר, יתחזק הטיעון בדבר השימוש ההוגן. מרבית מאגרי המידע שואפים לגיוון רב, על מנת למקסם את היתרונות ואת היקף השימושים (use-cases) הפוטנציאליים במערכת. לכן, במקרה הפרדיגמטי של בניית מאגרי מידע, קריטריון זה יצביע על הוגנות השימוש. במקרים בהם מאגר המידע אינו מגוון או מורכב מיצירות של יוצרים מעטים, יצביע שיקול זה נגד הוגנות השימוש.

פרשנות לפיה יצירה של מאגרי מידע לצורך למידת מכונה מהווה שימוש הוגן מתיישבת גם עם הרציונאלים המוכרים בספרות לדוקטרינת השימוש ההוגן. למעשה, הסוגיה דן מהווה מקרה פרדיגמטי להתקיימות כמה מן ההצדקות התאורטיות המרכזיות להגנת השימוש ההוגן. הצדקה מרכזית אחת לדוקטרינת השימוש ההוגן מעוגנת ביסוד הפרודוקטיביות. לפי תפיסה זו, דוקטרינת השימוש ההוגן מאפשרת שימושים רצויים שלא יכולים היו להתקיים בלעדיה.⁸⁴ הצדקה זו שלובה בהצדקה מרכזית נוספת לדוקטרינה זו, לפיה שימוש הוגן נועד להתמודד עם כשלי שוק, ובפרט עם עלויות עסקה גבוהות ועם "בעיית הסחטן", שישכלו שימושים רצויים ביצירות.⁸⁵ הצדקה נוספת מעוגנת בתפקידה של דוקטרינת השימוש ההוגן לייצר איזון בין זכות היוצרים לבין אינטרסים של משתמשים, שנדונה בהרחבה לעיל.⁸⁶

באשר לפרודוקטיביות, הרי שמתן אפשרות לעשות שימוש ביצירות יסיר מכשול מרכזי בפני למידת מכונה ויזניק קדימה את שוק הבינה המלאכותית ואת התעשיות שיכולות לעשות בו שימוש. ללא החלת הסדר השימוש ההוגן, יקשה על מיזמי בינה מלאכותית לצאת אל הפועל בשל כשלי שוק. ראשית, השגת רשיונות מבעלי זכויות היוצרים במליוני יצירות, שיכולים להמצא במקומות

⁸⁰ ראו למשל ע"א 8117/03 ענבר נ' יעקב, פס" 23 לפסק דינה של השופטת נאור (נבו 16.1.2006); ע"א (מחוזי ת"א) 3038/02 זום תקשורת (1992) בע"מ נ' הטלוויזיה החינוכית הישראלית, פס" 5(ו) לפסק הדין (נבו 29.4.2007), ת"א (שלום י-ם) 7036-09 רחמני נ' חברת החדשות הישראלית בע"מ (נבו 9.10.2011). ראו גם בירנהק, לעיל ה"ש 11; Lital; 8211-09 פורגס נ' בית חינוך תיכון גליל מערבי (נבו 27.7.2011).

⁸¹ ראו לעיל ה"ש 45-46 וטקסט נלווה.

⁸² בהקשר זה מעניין יהיה לעקוב אחר ההתפתחות בארצות הברית בעניין GitHub Copilot. ראו לעיל ה"ש 62.

⁸³ ראו סעיף 19(ב) לחוק, הקובע שרשימת השיקולים היא רשימה פתוחה.

⁸⁴ ראו דפנה לוינסון (זמיר), "ההגנה על שימוש הוגן בזכויות יוצרים" משפטים טז, 430, 432 (1985).

⁸⁵ ראו למשל Wendy J. Gordon & Daniel Bahls, *The Public's Right to Fair Use: Amending Section 107 to Avoid the "Fared Use" Fallacy*, 2007 UTAH L. REV. 619 (2007) (מפרטים רציונאלים לדוקטרינת השימוש ההוגן שמקורם בכשלי שוק).

⁸⁶ ש.ס. ראו גם אורית פישמן אפורי, היצירה הנגזרת בדיני זכויות יוצרים 331 (2005).

מדינת ישראל משרד המשפטים

שונים על הגלובוס מגלמת עלויות עסקה גבוהות באופן שמעמיד בספק את כדאיותם של מיזמי בינה מלאכותית מעיקרא.⁸⁷ על מנת שמיזמי בינה מלאכותית יוכלו לבנות מאגר מידע אפקטיבי, עליהם לאתר ולנהל משאים ומתנים עם כל אחד מבעלי זכויות אלה. על מנת לעמוד בתנאי הזכות המוסרית (הגם שלעמדתנו כאמור, אין הדבר נדרש), עליהם לאתר גם את היוצר עצמו, שאינו בהכרח בעל זכות היוצרים.⁸⁸ זאת ועוד. חלק מן היצירות "מקפלות" בתוכן זכויות אחדות. כך, למשל, אם יבקשו המיזמים לייצר מאגר מידע של קבצי מוסיקה, יהיה עליהם לנהל משא ומתן עם בעלי זכויות היוצרים במילים, בלחנים ובהקלטות ואולי גם עם מבצעי השירים ועם בעלי הזכויות בשידורים, הגם שדיון במגבלות על שימוש ביצירות מכוח זכויות של גורמים אלה לא נדון בחוות דעת זו.⁸⁹ ודוק. עלויות האיתור של בעלי הזכויות הרבים יוטלו רובן ככולן על היזמים. הדין לא מטיל כל חובה על בעל זכות יוצרים שיצירתו מועמדת לרשות הציבור לפרט את זהותם של בעלי הזכויות ביצירה, וחובה כאמור אינה מוטלת גם על הפלטפורמות בהן מועמדת היצירה לרשות הציבור. בפועל, רבות מן היצירות הניתנות לגישה באמצעות רשת האינטרנט אינן מפרטות את זהותם של בעלי הזכויות, והחוק אינו מקנה הגנה למשתמשים שמסתמכים על רישום מוטעה של זהות היוצר.⁹⁰

החשש לעלויות עסקה מסכלות שלוב בעניין זה בכשל שוק נוסף, הידוע כ"בעיית הסחטן" (holdup problem): פרגמנטציה של זכויות אקסקלוסיביות בין מספר עצום של בעלי זכויות נותנת בידיו של כל אחד מבעלי הזכויות הרבים כוח לסכל את הפרויקט, או למצער לעכב אותו באופן שהלכה למעשה ישים אותו לאל.⁹¹ אמנם כל יצירה במאגר המידע היא תחליפית וערכה של כל יצירה אינדוידואלית במאגר הוא שולי, ומשכך, באופן תאורטי, אין בכוחו של בעל זכויות בודד לסכל את הפרויקט כולו. אולם כשמובאים בחשבון התחרות בעולם היזמות והצורך לעמוד באבני דרך ביצועיות לצורך גיוס כספים ועמידה ביעדים, עיכוב הפרויקט עשוי להיות שקול לסיכולו.⁹² ודוק. העובדה כי מדובר בכשלי שוק משמעה כי **אי החלת ההגנות בחוק זכות יוצרים לא תביא לתשלום לבעלי זכויות יוצרים, אלא לכך שמיזמי בינה מלאכותית לא יצאו אל הפועל.**

בטרם נסכם סוגיה זו, ראוי להוסיף מספר מילים. קונספטואלית, הגנת השימוש ההוגן בוחנת שימושים לא מורשים ביצירות אד-הוק ולא באופן קטגורי. סוגיית השימוש ההוגן מוכרעת לרוב בדיעבד כאשר התבצע שימוש לא מורשה בתוכן מוגן והמשתמש מעלה את טענת השימוש ההוגן במסגרת הליך ליטיגציה (או פרה-ליטיגציה) אקס-פוסט. הטענה נבחנת אז אד-הוק, על בסיס ניתוח השימוש הקונקרטי שנעשה. לאור זאת, קביעה אקס-אנטה כי קטגוריה שלמה של שימושים חוסה תחת הגנת השימוש ההוגן עשויה להיראות אנומלית במידת מה.

על אף האמור, הניתוח המוצע כאן, לפיו בניית מאגרי נתונים לצורך למידת מכונה מהווה במרבית המקרים שימוש הוגן אין בה משום חריגה מגבולות הדוקטרינה. אדרבה, כאשר קיים מכנה משותף דומיננטי לקטגוריה מוגדרת של שימושים, בכוחה של פרשנות אקס-אנטה להסדר השימוש ההוגן לייצר וודאות בשוק עבור שחקנים משני צידי המתרס, לחסוך משאבי ליטיגציה ולקדם יעילות וודאות.⁹³

Wendy J. Gordon, *Fair Use as Market Failure: A Structural and Economic Analysis of the Betamax Case and Its Predecessors*, 82 COLUM. L. REV. 1600 (1982).

⁸⁸ ראו לעיל ה"ש 41 וטקסט נלווה.

⁸⁹ ראו ס' 4 לחוק זכות יוצרים; חוק זכות מבצעים ומשדרים, תשמ"ד-1984. דיון בזכויות מבצעים ומשדרים חורג מדין בחוות דעת זו.

⁹⁰ הגנת מפר תמים, הקבועה בסעיף 58 לחוק זכות יוצרים, מגנה רק על משתמשים שלא ידעו או שלא יכולים היו לדעת שיש זכויות יוצרים ביצירה.

⁹¹ Ben Depoorter & Francesco Parisi, *Fair Use and Copyright Protection: A Price Theory Explanation*, 21 INT'L REV. L. & ECON. 453 (2002). כך, למשל, לנוכח פסיקה שקבעה כי לכל אחד מאלפי אתלטים-לשעבר יש זכות למנוע יצירת משחק וידאו של פוטבול (Keller v. Elec. Arts, Inc., 724 F.3d 1268, 1269 (9th Cir. 2013)), ביכרה היצרנית הפוטנציאלית של המשחק לזנוח את הפרויקט על פני כניסה למשאים ומתנים אינדוידואליים.

⁹² ראו Lital Helman, *Innovation Policy and the Valley of Death*, Lital Helman, *Innovation Funding and the Valley of Death*, SMU L. REV., Forthcoming 2023, Available at <https://ssrn.com/abstract=4262740>.

⁹³ דוגמה להחלה של הגנת השימוש ההוגן בדין הישראלי על קטגוריה של שימושים היא סוגיית השימוש הפרטי ביצירות. בחוק הישראלי לא קיימת הגנה ספציפית לשימושים פרטיים ביצירות. ישנם גם מצבים בהם שימוש פרטי לא יחשב שימוש הוגן. ראו גם חוק זכויות יוצרים ומבצעים (קלטות) (תיקוני חקיקה), התשנ"ו-1996, הקובע מנגנון פיצוי בגין שימושים פרטיים בתכנים מוגנים בהקשרים מסויימים (ההסדר מוכר כ"חוק הקלטות הריקות"). לצד הסדר זה, מקובל גם בארה"ב לראות שימושים פרטיים כחוסים תחת הגנת השימוש ההוגן. ראו למשל Sony Corp. of

מדינת ישראל משרד המשפטים

עם זאת, אין בפרשנות המוצעת כאן כדי ליטול מהגנת השימוש ההוגן את עוקצה הקונקרטי הצמוד לעובדות המקרה. על אף הניתוח דכאן, יתכנו מקרים בהם ניתוח אד-הוק של שימוש מסויים ביצירות מוגנות לצורך למידת מכונה לא יחסה תחת הגנת השימוש ההוגן. טול למשל מקרה היפותטי בו חברה תרכיב מאגר נתונים מיצירות של צלם מסויים לשם אימון מכונה שתייצר יצירות בסגנון זהה לשל אותו צלם ותמכור אותן במחיר נמוך יותר או תתחרה בצלם במכרז על פרוייקט צילום גדול; או דמה בנפשך מכונה שמייצרת תקצירים של ספרי לימוד לאחר שאומנה על הטקסט המלא, ואשר פוגעת בשוק של ספרי הלימוד המקוריים. אף כי אין לקבוע מסמרות בעניינים אלה בהעדר מקרה ספציפי העומד לפנינו, הכרעה לפיה מקרים כגון אלה לא יחסו תחת כנפי ההגנה של שימוש הוגן תהא גם תהא קונסיסטנטית עם חוות דעת זו.

מעבר לעובדה שמקרים ספציפיים עשויים לחרוג מגבולותיה של דוקטרינת השימוש ההוגן, גם הטכנולוגיות המשמשות לבינה מלאכותית והשווקים הסובבים אותן עשויים לעבור טרנספורמציה מסוגים שונים. טכנולוגיות חדשות של למידת מכונה עשויות לא להזדקק להעתקה כלל או להתפתח לכיוונים שידרשו שימושים אחרים ביצירות. מערכות שיאפשרו מיקרו-תשלומים או מונטיזציה (הפקת הכנסה) אחרת של כלילת יצירות במאגרי מידע עשויות לשנות את פני התעשייה. אכן, התחום של למידת מכונה מתפתח בקצב מהיר והסוגיות שהוא מעורר הן דינאמיות ורבגוניות. חוות דעת זו אינה מבקשת לקבוע את המצב הקיים. אדרבה, מכירה היא בדינאמיות האינהרנטית לתחום זה ומותירה במתכוון שוליים גמישים בהחלת דוקטרינת השימוש ההוגן על למידת מכונה.

ב. שימוש אגבי ביצירה

הסדר נוסף שעשוי לחול בהקשר דנן עניינו שימוש אגבי ביצירה. סעיף 22 לחוק מתיר שימוש אגבי ביצירה בדרך של הכללתה ביצירת צילום, ביצירה קולנועית או בתקליט, כך:

"שימוש אגבי ביצירה בדרך של הכללתה ביצירת צילום, ביצירה קולנועית או בתקליט, וכן שימוש ביצירה שבה הוכללה היצירה כאמור, מותר; לעניין זה, הכללה במתכוון של יצירה מוסיקלית, לרבות המילים הנלוות אליה, או של תקליט שבו היא טבועה, ביצירה אחרת, לא תיחשב לשימוש אגבי."

על מנת להחיל את הסעיף בהקשר דנן, יש לראות את תהליך למידת המכונה כשימוש אגבי ביצירות על מנת לייצר את התוצר של מערכת הבינה המלאכותית. כאשר התוצר של המכונה הוא יצירת צילום, יצירה קולנועית או תקליט, ניתן יהיה לראות בהעתקת יצירות למאגר הנתונים לצורך למידת המכונה משום שימוש אגבי. זאת משום שריבוי היצירות המרכיבות את המאגר משמעו כי כל יצירה אינדיווידואלית במאגר הנתונים מהווה רכיב זניח בלמידה ואין בה משום תוספת משמעותית למאגר הנתונים או לפונקציונאליות של המערכת.⁹⁴ יתר על כן, התוצר הסופי של המכונה לא מכיל את היצירה כלל. במובן זה, "אגביות" השימוש מוגברת בהקשר הנוכחי לעומת כוונת המחוקק המקורית: ההנחה בסעיף 22 היא שהיצירה המוגנת נכללת - הגם שבאופן לא מטריאלי - ביצירה החדשה.⁹⁵ בהקשר דנן, היצירה לא תכלל אלא במאגר הנתונים, שמהווה אך חומר גלם ללמידת המכונה, ולא תשתקף כלל ביצירה הסופית.

בעניין זה ניתן להסתייע בפסיקה האמריקאית. שם, בפרשת **Baker**, נקבע כי שימוש בתכנים מוגנים יהא מותר כאשר הוא אינצינדנטלי והכרחי לשימוש לא מפר.⁹⁶ פסיקה זו יושמה בהקשרים

America v. Universal City Studios, Inc, 464 U.S. 417 (1984) (פוסק כי שימוש פרטי לא מסחרי מוחזק כהוגן); Frances Grodzinsky & Maria C. Bottis, *Private Use as Fair Use: Is it Fair?*, 37(2) COMP. & SOC. 11, 12 (2007) ("Private use in US law is a sub-set of fair use.")

⁹⁴ גרינמן, לעיל ה"ש 12, עמ' 340-343. ראו גם ת"א 53689-10-17 ברדוגו נ' ד. איתן / ר. להב-ריג אדריכלים ומתכנני ערים בע"מ ואח', פסי' 55 לפסק הדין (16.8.20).

⁹⁵ ראו דברי ההסבר להצעת חוק זכות יוצרים: "...עקרונות עלולה להיות הפרת זכות יוצרים במוסיקה הנשמעת ברקע בעת צילום אירוע ציבורי, או בעת צילום אירוע ציבורי או בתמונה התלויה על הקיר במקום שבו מצלמים סרט או כתבה חדשותית. אף שמעשים אלה עשויים להיחשב "כזוטי דברים", מוצע להבהיר כי אינם מהווים הפרת זכות יוצרים" (ה"ח תשס"ה מס' 196, עמ' 1116).

⁹⁶ Baker v. Selden, 101 U.S. 99, 104 (1880).

מדינת ישראל משרד המשפטים

של שימוש במידע מוגן בתהליכי הנדסה חוזרת (reverse engineering) של תוכנה, והיא רלוונטית גם להקשר דנן, ככל שהתוצר של המערכת אינו מפר בעצמו את זכות היוצרים.⁹⁷

יש לסייג את תחולת ההסדר של שימוש אגבי בהקשר דנן בשני מובנים. ראשית, החוק מייחד הסדר זה ליצירות מסוג יצירת צילום, יצירה קולנועית או תקליט, ומחריג מצבים בהם היצירה הסופית כוללת במתכוון יצירה מוסיקלית או תקליט. סייגים אלה יחולו כמובן גם על למידת מכונה. משמעות הדבר היא כי הפרוייקטים שיוכלו להשתמש בחריג זה יכללו מערכות בהן התוצר של למידת המכונה יהא בעצמו יצירה מסוג יצירת צילום, יצירה קולנועית או תקליט, ולא, למשל, טקסט או לחן. באופן עקרוני, החריג הקבוע בסיפא של הסעיף, בדבר הכללה במתכוון של יצירה מוסיקלית או תקליט יחול כמובן גם הוא. יחד עם זאת, אין לפרש חריג אחרון זה בהרחבה יתרה. טכנית, כל היצירות במאגר נכללו בו במתכוון, שאם לא כן, היו מושמטות ממנו בהליך גיבוש המאגר. אולם עמדתנו היא כי כל עוד המאגר מגוון ולא נכללו בו במתכוון דווקא יצירות של יוצרים בודדים בלבד, עשויה לקום תחולה להסדר הנדון.

המובן השני בו יש לסייג את תחולת סעיף 22 עניינו מצבים עובדתיים בהם השימוש אינו אגבי. דוגמה אפשרית לכך היא מערכת שמתאמנת על מאגר מידע המורכב מיצירות של יוצר מסויים בלבד והתוצר שלה הינו יצירות דומות מסוג יצירת צילום, יצירה קולנועית או תקליט. דוגמה נוספת היא חברה שעושה שימוש מסחרי או שימוש מהותי אחר במאגר המידע עצמו או ביצירות הכלולות בו. לדוגמה, עבור חברה שהמודל העסקי שלה מבוסס על איסוף תכנים, "ניקוי" שלהם ויצירת מאגרי מידע עבור חברות שמפתחות מערכות מבוססות בינה-מלאכותית, השימוש ביצירות שמרכיבות את מאגר המידע לא יוכל במקרה הרגיל להיחשב לשימוש אגבי.

ג. העתקה זמנית

הסדר נוסף שעשוי לאפשר שימושים ביצירות מוגנות לצורך למידת מכונה רלוונטי למצבים בהם העתקת היצירות למאגר המידע היא זמנית. ההיתר לבצע העתקה זמנית של יצירות מוגנות מעוגן בסעיף 26 לחוק בזו הלשון:

"26. העתקה זמנית, לרבות העתקה כאמור שנעשתה בדרך אגבית, של יצירה, מותרת אם היא מהווה חלק בלתי נפרד מהליך טכנולוגי שמטרתו היחידה היא לאפשר העברה של היצירה בין צדדים ברשת תקשורת, על ידי גורם ביניים, או לאפשר שימוש אחר כדין ביצירה, ובלבד שאין לעותק האמור ערך כלכלי משמעותי משל עצמו."

מלשון החוק עולה כי העתקה זמנית תהווה שימוש מותר ביצירה במצבים המנויים בסעיף וכאשר לעותקים בהם נעשה השימוש אין ערך כלכלי עצמאי. מבין המצבים המנויים בסעיף, יצירת מאגר נתונים לשם למידת מכונה יחסה תחת הגנת הסעיף אם מטרתו היא "לאפשר שימוש כדין ביצירה".⁹⁸ תנאי זה יתקיים במצבים בהם מלבד העתקת התכנים המוגנים למאגר הנתונים, השימוש ביצירה נעשה "כדין", למשל, כאשר הגישה הראשונית ליצירות נעשתה כדין. מלבד תנאי זה, מותנית הגנת הסעיף בכך שלעותק היצירה אין ערך כלכלי משמעותי.⁹⁹ בעניין דנן, נראה כי תנאי זה מתקיים, שהרי לכל יצירה אינדיווידואלית במאגר הנתונים ערך מזערי. עם זאת, הרי למאגר הנתונים עצמו עשוי להיות ערך כלכלי משמעותי עד מאד, בפרט לאחר ש"נוקה" ו"תוויג" והותאם לקריאת מכונה. אם המאגר הוא בעל ערך עצמאי, הסדר השימוש הזמני לא יתאים כנראה להגנה עליו. ממילא, על מנת שמאגר הנתונים יוכל לחסות תחת הגנת סעיף 26, עליו להימחק בתום השימוש.

ברי כי הפרשנות לפיה סעיף 26 עשוי להגן על יצירת מאגרי מידע ללמידת מכונה שנוצרים באופן זמני היא פרשנות מרחיבה. במקור, נועד הסעיף להגן על תהליכים טכנולוגיים אוטומטיים שמייצרים העתקים של קטעי קוד או של יצירות מוגנות אחרות באופן בלתי נמנע אגב פעילות רגילה של תוכנה. אולם במועד כתיבת החוק, טכנולוגיות של למידת מכונה או מקבילות שלה לא עמדו

⁹⁷ Pamela Samuelson, *The Story of Baker v. Selden: Sharpening the Distinction between Authorship and Invention*, INTELLECTUAL PROPERTY STORIES 159-193 (2005).

⁹⁸ ראו ס' 26 לחוק זכות היוצרים.

⁹⁹ הפרשנות המקובלת למונח ערך כלכלי היא שגם במקרה שלעותק של היצירה יש ערך כלכלי פוטנציאלי, מי שאינו מממש ערך זה יהנה מהפטור. גרינמן, לעיל ה"ש 12, בעמ' 365.

מדינת ישראל

משרד המשפטים

בפני המחוקק, ואין לפרש את הסעיף כיוצר הסדר שלילי.¹⁰⁰ יתר על כן, גם האיחוד האירופי הכיר בסעיף 15(1) לדירקטיבה, הקובע חריג של שימוש זמני, ככזה היכול לחול על הסוגיה באופן עקרוני.¹⁰¹ חריג ההעתקה הזמנית על פי הדין האירופי מקביל לזה הישראלי, ויחול במצבים בהם העתקת התכנים היא זמנית (transient) או אינצידנטלית לתהליך טכנולוגי הנדרש לשימוש שאינו מפר ושאינו בעל ערך כלכלי עצמאי.¹⁰²

לסיכום, גם אם מאגר מידע זמני אינו המקרה התדיר של למידת מכוונה כיום, בשל המשאבים הרבים המושקעים בדרך כלל בתהליך יצירת המאגר, הרי שבמקרים בהם מאגר המידע נמחק בתום השימוש בו, ההסדר הקבוע בסעיף 26 יוכל להגן מפני טענת הפרה של זכות יוצרים. הסעיף לא יוכל להגן על מיזמים שישמרו את מאגר הנתונים וגם לא על חברות שייצרו מאגר נתונים עבור מיזמי בינה מלאכותית. שחקנים אלה מנצלים את הערך הכלכלי של מאגר הנתונים ואינם עושים בו שימוש זמני.

הפרשנות שהוצעה בפרק זה לחוק זכות יוצרים נועדה ליצור איזון בין ההבטחות הגלומות בלמידת מכוונה לבין זכויות היוצרים ביצירות המרכיבות את מאגר הנתונים. על פי הניתוח המוצע בפרק זה, דוקטרינת השימוש ההוגן תפרוש כנפיה על העתקת יצירות מוגנות לצורך יצירת מאגר נתונים ללמידת מכוונה. כמוה, הסדר השימוש האגבי והסדר ההעתקה הזמנית יספקו גם הם הגנה על פעילות זו במצבים מסויימים, כמפורט לעיל. פרשנות זו תאפשר פריחה של שוק הבינה המלאכותית תוך פגיעה מינימלית - אם בכלל - בבעלי זכויות יוצרים. ברם, כפי שכבר צויין, מצבים מסויימים יצדיקו סטייה מן הכלל האמור והסטה של נקודת האיזון לעבר הגנה מוגברת על זכויות היוצרים. חלק מן המצבים האמורים נדונים בחוות דעת זו וחלקם טמונים בחיק העתיד ואינם ברי צפיה בנקודת הזמן הנוכחית. הפרק הבא יפרט כיצד התמודדו מדינות אחרות ברחבי העולם עם הסוגיה.

ד. משפט משווה

מדינות העולם ערות לצורך באיזון בין זכויות יוצרים לבין האינטרסים הגלומים בשימושים מותרים ביצירות.¹⁰³ שיטות המשפט השונות מודעות היטב גם לאתגרים הייחודיים של המתח בין זכויות יוצרים לבינה מלאכותית בכלל ולשאלת מאגרי הנתונים ללמידת מכוונה בפרט.¹⁰⁴ כפי שנראה להלן, המגמה הרווחת ברחבי העולם עולה בקנה אחד עם חוות דעת זו, ומאפשרת שימוש ביצירות מוגנות לצורך בנייה של מאגרי מידע ללמידת מכוונה, ברמות משתנות של גמישות.

חלק משיטות המשפט בחרו להתמודד עם הסוגיה באמצעות יצירת הסדר ספציפי בנוגע לשימוש ביצירות לצורך למידת מכוונה. בשיטות משפט אלה, נקבע ההסדר במסגרת חריג ה-TDM.

¹⁰⁰ חידושים טכנולוגיים מחייבים תדיר בחינה מחודשת של דיני זכויות היוצרים ושל האיזונים הכרוכים בהם. ראו למשל בש"א (מחוזי-ת"א) 11646/08 The Football Association Premier League Ltd נ' פלוני, פס" 1 לפסק הדין (נבו) 2.9.2009, נהפך בעליון בע"א 9183/09, לעיל ה"ש 4 ("המהפכה הדיגיטלית ורשת האינטרנט שינו את חיינו. [...]) שינוי זה מחייב איזונים חדשים שייקחו בחשבון את זכויות הציבור בצריכת תרבות והשתתפות בשיח התרבותי..."); עניין טלראן, לעיל ה"ש 9 (נדרש לשאלה אם מכירת מכשיר המאפשר "עקיפה" של אמצעי הגנה טכנולוגיים מהווה הפרת זכות יוצרים).

¹⁰¹ ראו ס' 15(1) לדירקטיבת Infococ; ראו גם European Parliament, לעיל ה"ש 46, בעמ' 11. ¹⁰² ראו, CJEU, C-360-13, Public Relations Consultants Association (5 June 2014), ECLI: EU: C: 2014: 1195, §§ 43, 50 ("[an act of reproduction is incidental] if it neither exists independently of, nor has a purpose independent of, the technological process of which it forms part").

¹⁰³ על אף תמימות הדעים באשר לצורך באיזון בין זכויות יוצרים לאינטרסים הגלומים בגישה לתכנים, נחלקות מדינות העולם בין גישות שונות באשר למנגנוני האיזון הרצויים. מנגנון איזון נפוץ הינו מנגנון הטיפול ההוגן (fair dealing) שמקורו בבריטניה והוא נפוץ במרבית מדינות העולם. מנגנון זה יוצר איזון באמצעות קביעת רשימה סגורה של שימושים מותרים ביצירות מוגנות (אם כי לעתים חריגים אלה מפורשים בהרחבה). גישה דומיננטית נוספת היא גישת השימוש ההוגן (fair use) שמקורה בארצות הברית והיא התפשטה לישראל, סינגפור, פולין, הפיליפינים ודרום קוריא. דוקטרינת השימוש ההוגן מרחיבה את קשת השימושים המותרים ביצירות מוגנות ומספקת מבחנים להגנות השימוש, חלף רשימה סגורה של שימושים מותרים. שיטות משפט אחרות קובעות מנגנוני איזון אחרים, כגון חריגים סטטוטוריים ספציפיים בלבד לזכויות היוצרים, בלי מנגנון של שימוש הוגן או טיפול הוגן. גם שיטות משפט בהן קיימים מנגנונים של שימוש הוגן או טיפול הוגן יכולות לכלול לצידם חריגים סטטוטוריים מוגדרים.

¹⁰⁴ ראו למשל, OECD Council, Recommendation of the Council on Artificial Intelligence, available at <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449> (Adopted on May 22, 2019; 2021), OECD/LEGAL/0449.

מדינת ישראל משרד המשפטים

Text & Data Mining - חריג ה-TDM נוצר לפני עידן למידת המכונה, והוא נועד בבסיסו להסדיר איסוף וניתוח אוטומטיים של מידע במרחב הדיגיטלי. עד עתה, בריטניה, יפן, סינגפור והאיחוד האירופי נקטו באסטרטגיה זו, והחילו את חריג ה-TDM על יצירת מאגרי מידע ללמידת מכונה. קנדה ודרום קוריאה שוקלות לפסוע בדרך זו גם הן.¹⁰⁵ לעומתן, ארצות הברית, המובילה העולמית היום בפיתוח טכנולוגיות ושווקים ללמידת מכונה, נמנעת מיצירת חריג סטטוטורי, ונסמכת על הגנת השימוש ההוגן על מנת לייצר את האיזון הרצוי, כפי שמוצע לעשות בחוות דעת זו. להלן יפורטו בקצרה ההסדרים הקיימים בשיטות המשפט שנתנו את דעתן לסוגיה וידונו השיקולים לבחירה בין האסטרטגיות האמורות. כפי שידון להלן, יתרונה של האסטרטגיה הראשונה הוא ביצירת וודאות ויתרונה של השנייה הוא בגמישותה וביכולת ההתאמה שלה להתפתחויות טכנולוגיות, משפטיות וכלכליות.

החלוצה בקביעת הסדר ספציפי לסוגיית השימושים המותרים לצורך לימוד מכונה היתה יפן. כמרבית חוקי זכויות היוצרים ברחבי העולם, החוק היפני אינו כולל הסדר של שימוש הוגן. חלף זאת, כולל הוא רשימה סגורה של שימושים מותרים ביצירות המנויים בחוק.¹⁰⁶ רשימה זו כוללת בין היתר שימוש ביצירות במסגרת נאומים פוליטיים, תצוגה של יצירות אמנות על ידי בעליהן, התאמת יצירה לטובת לקויי-שמיעה ושימושים רבים נוספים.¹⁰⁷ בשנת 2018, היתה יפן למדינה הראשונה שתיקנה את חוק זכות היוצרים שלה והוסיפה בו סעיף המתיר ניצול יצירה לצורך TDM ללא צורך ברשיון, ללא תשלום לבעל הזכויות וללא צורך במתן קרדיט ליוצר.¹⁰⁸ הסייג היחיד להיתר זה עניינו מצבים בהם ניצול היצירה פוגע באופן בלתי סביר באינטרס כלכלי של היוצר לאור טיבה או תכליתה של היצירה או לאור נסיבות ניצולה.¹⁰⁹

שנה לאחר מכן, בשנת 2019, התקבלה באיחוד האירופי הדירקטיבה המוכרת כ-DSM, ונקבעו בה שני הסדרים המתירים שימוש ביצירות מוגנות לצורך TDM.¹¹⁰ ההסדר הראשון, המעוגן בסעיף 3 לדירקטיבה, מאפשר העתקה של יצירות מוגנות על ידי מוסדות מחקר או מורשת לצורך מחקר מדעי.¹¹¹ הסדר זה טרם אומץ בכל מדינות האיחוד ולעתים אומץ בצמצום.¹¹² ההסדר השני, המופיע בסעיף 4 לדירקטיבה, קובע כי כל אדם יהא רשאי לבצע TDM ביצירות מוגנות, לכל שימוש שהוא, ובלבד שהיצירות המוגנות ימחקו לאחר איסוף המידע ושההסדר יכלול מנגנון שיאפשר לבעלי זכויות המתנגדים לו להחריג את יצירותיהם מההסדר (out opt).¹¹³ יוזכר כי כפי

¹⁰⁵ ראו למשל Seung Hoon Park, *Copyright Issues for AI and Deep Learning Services: A Comparison of U.S., South Korean, and Japanese Law*, JTIP BLOG (May 28, 2021) <https://jtjp.law.northwestern.edu/2021/05/28/copyright-issues-for-ai-and-deep-learning-services-a-comparison-of-u-s-south-korean-and-japanese-law/> (מצייין כי דרום קוריאה שוקלת חריג כאמור).
¹⁰⁶ Chosakuken Ha [Japanese Copyright Act], Law No. 48 of 1970, art. 30-4, translated in CHOSAKUKEN KANKEI HOREI DETA BESU [COPYRIGHT-RELATED LAW DATABASE] (Copyright Research and Information Center (CRIC)), available at <https://www.cric.or.jp/english/clj/cl2.html> (להלן: "חוק זכות היוצרים היפני").

¹⁰⁷ שם. שימושים אלה מקבילים בחלקם הגדול לרשימת השימושים המותרים הקבועים בחוק הישראלי לצד הסדר השימוש ההוגן השווה לסי' 18-32 לחוק זכות יוצרים.
¹⁰⁸ ראו סעיף 4-30 לחוק זכות היוצרים היפני, לעיל ה"ש 106.

¹⁰⁹ שם.
¹¹⁰ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, art. 3-4, 2019 O.J. (L 130) 92 (EU) (להלן: "דירקטיבת DSM"). ראו גם Joao Pedro Quintais (2017), *Rethinking Normal Exploitation: Enabling Online Limitations in EU Copyright Law*, AMI No. 6, pp. 197-205.
¹¹¹ ראו סי' 3 לדירקטיבת DSM.

¹¹² במספר מדינות, כגון פינלנד, ההסדר טרם אומץ נכון לכתיבת שורות אלה. ראו Sofia Wang, *Implementation of the DSM Directive is progressing in Finland — what changes will the directive bring to the Finnish copyright legislation?*, BIRD & BIRD (June 20, 2022), available at <https://www.twobirds.com/en/insights/2022/finland/implementation-of-the-dsm-directive-is-progressing-in-finland-what-changes-will-the-directive-bring-to-the-finnish-copyright-legislation/>. באשר לאימוץ גרסה מצמצמת של החריג, ראו למשל החוק הצרפתי - Art. 10 added paragraph 10 to Art. L122-5 and paragraph 5 to Art. L342-3 of the Intellectual Property Code (Code de la propriété intellectuelle) (CPI); החוק האסטוני - Estonian Copyright Act, Art. 19(3); החוק הגרמני הקדים את הדירקטיבה בשנתיים וכלל חריג TDM בהקשר לא-מסחרי בלבד. ראו Act on Copyright and Related Rights (Urheberrechtsgesetz, UrhG) as last amended by Article 1 of the Act of 1 September 2017 (Federal Law Gazette I p. 3346), The Federal Ministry of Justice and Consumer Protection, translation by Ute Reusch, section 60d, available at http://www.gesetze-im-internet.de/englisch_urhg. עם כניסת הדירקטיבה לתוקף שונה החוק הגרמני ב-2021 כדי לכלול אותה באופן מלא. מדינות אחרות, כגון הולנד, הטמיעו את חריגי הדירקטיבה באופן מלא. ראו: Annemarie Beunen, *Copyright directive series - a closer look at the Netherlands*, europeana pro (august 3, 2022), available at <https://pro.europeana.eu/post/copyright-directive-series-a-closer-look-at-the-netherlands>.
¹¹³ ראו סי' 4 לדירקטיבת DSM.

מדינת ישראל משרד המשפטים

שצויין לעיל, האיחוד האירופי קבע בנוסף לחריגי ה-TDM גם חריג בדבר שימוש חולף או אינצידנטלי ביצירות, שעשוי גם הוא להיות שימושי בהקשר דנן.¹¹⁴

שלישית להצטרף לאסטרטגיית ההסדרים הייעודיים היתה **סינגפור**, בסוף שנת 2021.¹¹⁵ אימוץ אסטרטגיה זו על ידי סינגפור היא מעניינת מכיוון שבניגוד לחברות האחרות ב"מועדון" חריגי ה-TDM, חוק זכות היוצרים בסינגפור דווקא כולל הסדר של שימוש הוגן. סעיף השימוש ההוגן הסינגפורי דומה עד מאד לזה הישראלי ולזה האמריקאי, ויכול היה לשימש מסגרת לקביעת האיזונים בסוגיה זו. על אף האמור, הכריעה סינגפור לטובת יצירת חריג סטטוטורי עבור יצירת מאגרי מידע לצורך למידת מכונה, וזאת על מנת להגביר את הוודאות עבור תעשיית הבינה המלאכותית ולבצר בכך את תכניתה להפוך את סינגפור ל"האב" (hub) גלובלי לבינה מלאכותית.¹¹⁶ חריג ה-TDM בסינגפור הוא רחב ביותר, וכולל, בין היתר, תחולה בהקשר מסחרי, רשימה רחבה ופתוחה של פעולות המוגדרות כ-TDM והתגברות על תניות חוזיות העלולות לסכל TDM.¹¹⁷

בשנת 2022, הרחיבה **בריטניה** את חריג ה-TDM שנחקק על ידה לראשונה בשנת 2014. חריג ה-TDM המקורי בבריטניה, כבמדינות רבות באירופה, הוחל רק בהקשר מחקרי שאינו מסחרי.¹¹⁸ לאחר יצירת תכנית לאומית לבינה מלאכותית בבריטניה (ויתכן שלא במנותק ממנה) הושמטה מהחוק ההגבלה האמורה על חריג ה-TDM, והחוק הבריטי מאפשר עתה שימוש ביצירות מוגנות לצורך יצירת מאגרי מידע ללמידת מכונה גם בהקשר מסחרי.¹¹⁹

מדינות נוספות שוקלות לקבוע בחקיקה חריג לדיני זכויות יוצרים בהקשר של למידת מכונה. הדין **הקנדי**, למשל אינו כולל הסדר של שימוש הוגן, אך כולל מנגנון דומה הנקרא טיפול הוגן (fair dealing).¹²⁰ עד לכתיבת שורות אלה, לא נדונה בפסיקה הקנדית סוגיית התחולה של הסדר הטיפול ההוגן על למידת מכונה.¹²¹ ב-2019, לאור שאיפתה של קנדה לקדם בתחומה את תחום הבינה המלאכותית,¹²² המליצה הועדה הפרלמנטרית שעסקה בחוק זכות יוצרים להתיר בחקיקה שימוש חופשי ביצירות מוגנות לצורך "ניתוח מידע" (informational analysis) ללא אישור היוצר או להוסיף "ניתוח מידע" כקטגוריה נוספת בסעיף הטיפול ההוגן.¹²³ המלצות הועדה האמורה, ובכללן המלצה זו, טרם יושמו.

שיטת משפט שנמנעת באופן מובהק מאימוץ חריג סטטוטורי ספציפי ללמידת מכונה היא השיטה **האמריקאית**. למעשה, הפרשנות המוצעת בחוות דעת זו, לפיה שימוש בתכנים מוגנים לצורך למידת מכונה מהווה לרוב שימוש הוגן, נקבעה קודם לכן בדין האמריקאי, ממנו אימץ כאמור המשפט הישראלי את דוקטרינת השימוש ההוגן.¹²⁴ הכרעה זו נקבעה בפרשת Authors Guild v. Google (פרשת "Google Books"), בה פסק בית המשפט כי מיזם הספרים של גוגל,

¹¹⁴ ראו לעיל ה"ש 101 וטקסט מלווה.
¹¹⁵ ראו ס' 243 לחוק זכות יוצרים הסינגפורי (ניתן לצפיה ב: <https://sso.agc.gov.sg/Acts-Supp/22-2021/Published/ProvIds=P15-#pr243>).

¹¹⁶ ראו ממשלת סינגפור, תכנית לאומית לבינה מלאכותית (ניתנת לצפיה ב: vision: "By 2030, Singapore will be a leader in developing and deploying scalable, impactful AI solutions, in key sectors of high value and relevance to our citizens and businesses." (https://www.smartnation.gov.sg/initiatives/artificial-intelligence).

¹¹⁷ ראו ס' 244 לחוק זכות יוצרים הסינגפורי.
¹¹⁸ Regulation 3 of the Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014, No. 1372, adding Article 29A to the Copyright, Designs and Patents Act 1988. The Regulations came into force on 1 June 2014.

¹¹⁹ ראו ס' 29A לחוק זכות היוצרים הבריטי.
¹²⁰ Canada Copyright Law, R.S.C. 1985, c C-42, art. 29. בפועל, פרשנות מרחיבה לחריג ה"טיפול ההוגן" במהלך השנים מעממה את מרבית ההבדלים בינו לבין סעיף השימוש ההוגן האמריקאי. פסק הדין המוביל בעניין הקריטריונים לטיפול הוגן בקנדה הוא CCH Canadian Ltd. v. Law Society of Upper Canada [2004] S.C.R. 339, 342 (Can.).
¹²¹ Brown, Byl & Grossman, "fair dealing", *Canada Is*; 161-159, בעמ' 54, לעיל ה"ש 54, בעמ' 159-161, Christopher Guly, *Gathering Public Input on Copyright Implications of AI, Internet of Things* (august 13, 2021) ("Given the absence of a clear rule to permit machine learning in Canadian copyright law (often called a text and data mining exception), our legal framework trails behind other countries that have reduced risks associated with using data sets in AI activities") available at <https://www.cigionline.org/articles/canada-is-gathering-public-input-on-copyright-implications-of-ai-internet-of-things/>.

¹²² ראו <https://ised-isde.canada.ca/site/ai-strategy/en> (last visited: 8.28.2022).

¹²³ DAN RUIFY, HOUSE OF COMMONS, STATUTORY REVIEW OF THE COPYRIGHTS ACT: REPORT OF THE STANDING COMMITTEE ON INDUSTRY, SCIENCE AND TECHNOLOGY 85-87 (June 2019).

¹²⁴ ראו למשל 657 (2016) IOWA L. REV. 657, James Grimmelman, *Copyright for Literate Robots*, 101 IOWA L. REV. 657 (2016); Benjamin L.W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J.L. & ARTS 45, 96 (2017).

מדינת ישראל משרד המשפטים

במסגרתו סרקה החברה לשרתיה מליוני ספרים במטרה לשמרם ולהנגישם לציבור, חוסה תחת הגנת השימוש ההוגן.¹²⁵

במהלך השנים פורשה פסיקה זו בהרחבה ניכרת, ככזו המאפשרת כל פעולה ביצירה מוגנת שהקהל שלה הוא מחשב ולא אדם. כפי שתאר זאת המלומד גיימס גרימלמן -

“[Q]uietly, invisibly, almost by accident, copyright has concluded
¹²⁶ that reading by robots doesn't count.”

בשנים האחרונות, נראה כי חל מפנה מסויים בפרשנות הרחבה לפיה **“reading by robots doesn't count”**. שני פסקי דין חדשים עשויים להתפרש כמעמעמים את הוודאות המשפטית דה-פקטו שפעולות אוטומטיות נהנו מהן עד עתה. פסקי דין אלה דחו לאחרונה טענות לפיהן פעולות שביצעה מערכת מחשב ביצירות מוגנות מהוות שימוש הוגן. באחד מפסקי הדין נדונה פלטפורמה שביצעה העתקה של תכנים חדשתיים ואפשרה למשתמשים לערוך בהם חיפוש, ובשני נדון שירות שאיפשר חיפוש, צפיה והעתקה של שידורים מוגנים.¹²⁷

ספק אם מהלכים פסיקתיים זה ישפיעו באופן כלשהו על שוק הבינה המלאכותית ועל הסוגיה של למידת מכונה. פסקי הדין הללו אינם עוסקים במערכות לומדות או בבינה מלאכותית, ובהחלט סביר שתקדים Google Books יוסיף לחול באין מפריע בסיטואציה של למידת מכונה שבה נוצר. אכן, ההבדל בין מיזמי למידת מכונה לבין הדין החדשים הוא משמעותי. בשני פסקי הדין החדשים, השימוש שביצעה המערכת ביצירות כלל הצגה של חלק גדול מן התכנים המקוריים למשתמשי קצה ללא שינוי ובכך פגע בשוק המקורי של בעלי הזכויות, שביקשו לספק גישת משתמשים לתכנים אלה בעצמם. במובחן מכך, השימושים הסטנדרטיים של מערכות למידת מכונה לא מקנות גישה ליצירות הכלולות במאגר המידע שלא לצורך למידת מכונה. על אף האמור, למען הזהירות יצוין כי פסקי הדין הללו עשויים אולי להצביע על מיתון מדיניות היד החופשית כלפי פעילותן של מערכות אוטומטיות, ולהשפיע בשלב כלשהו גם על הדין שיחול ביחס ללמידת מכונה.

לסיכום, נראה כי מדינות רבות על פני הגלובוס מכירות בכך שללא מתן אפשרות לעשות שימוש אוטונומי ביצירות מוגנות, הן לא יוכלו להגשים את שאיפותיהן בתחום של בינה מלאכותית. לאור זאת, **מדינות שנתנו את הדעת לסוגיה קבעו כי יש להכיר בחריגים לצורך זה**, הגם שכל שיטת משפט קבעה באופן אחר את היקף ההגנה ואת האופן בו היא תינתן, אם באמצעות הסדרים יעודיים או באמצעות פרשנות לדוקטרינות קיימות.

היתרון המובהק של גישת ההסדרים הייעודיים נעוץ ביצירת וודאות בשוקי היצירה והבינה המלאכותית. אולם יתרון זה קצר ימים הוא ועשוי ברבות הימים להפוך לחסרון של ממש. חריג ה-TDM מהווה דוגמה מובהקת לעניין. עד לפני כעשור, הסוגיות שהעסיקו משפטנים בתחום של שימושים אוטומטיים ביצירות סבבו סביב סוגיות של איסוף וניתוח מידע. זו הסיבה שמושג ה-TDM היה - ונותר - הפרדיגמה לקביעת חריגים בתחום של למידת מכונה. ואולם, המסגרת של TDM, שמאפיינת כאמור את הסוגיה ככזו שעוסקת בכריית המידע ובאיסופו ולא ביצירת מאגר נתונים דינאמי ובהסקת מסקנות ממנו, מותירה מספר סוגיות בלתי פתורות בהקשר של טכנולוגיות עדכניות ללמידת מכונה. כך, למשל, הנחת העבודה היא שהמידע אותו מבקשים לאסוף אינו מוגן כשלעצמו, כגון נתונים, ורק הדרך אליו עוברת בהעתקה של חומר מוגן.¹²⁸ על בסיס הנחה זו, אחד מחריגי ה-TDM באיחוד האירופי מחייב מחיקת היצירות לאחר איסוף המידע, כפי שנדון לעיל. כמפורט בחוות דעת זו, לאור התשומות המושקעות ביצירת מאגרי מידע והשימושים הדינאמיים במאגרים לאורך חיי המיזם, דרישה זו פוגעת ברלוונטיות של הסעיף למיזמים אקטואליים של למידת מכונה, הנזקקים למאגרי המידע שיצרו גם בהמשך.¹²⁹ אמנם חלק מהמדינות הגדירו את

¹²⁵ **שם**. ראו גם Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 2014 U.S. App. LEXIS 10803, Copy. L. Rep. (United States Court of Appeals for the Second Circuit June 10, 2014, Decided). (CCH) P30,617, 42 Media L. Rep. 1898, 111 U.S.P.Q.2D (BNA) 1001, 2014 WL 2576342

¹²⁶ ראו Grimmelmann, **לעיל** ה"ש 124, שם.
¹²⁷ Fox News Network, LLC v. TVEyes, Inc., 883 F. 3d 169 (2nd Cir. 2018); Associated Press v. Meltwater, 128 F.3d 1171 (2d Cir. 2003); U.S. Holdings, Inc., 931 F. Supp. 2d 537, 543-44 (S.D.N.Y. 2013). **לעיל** ה"ש 60.

¹²⁸ ראו גם Gervais, **לעיל** ה"ש 69, שם.
¹²⁹ ראו גם European Parliament, **לעיל** ה"ש 46, בעמ' 12 (“In sum, existing exceptions and limitations might not offer a stable legal framework to safely engage in TDM research projects and invest considerable resources.”).

מדינת ישראל

משרד המשפטים

מושג ה-TDM בהרחבה וכללו בו גם פונקציות עדכניות של מערכות לומדות. אולם גם הגדרות מודרניות מעוגנות בסופו של יום ביכולות הטכנולוגיות הנוכחיות. יכולתם של הסדרים ספציפיים לצפות פני עתיד היא מטבע הדברים מוגבלת, והמחיר הנגבה עבור הוודאות הוא גבוה. טכנולוגיות מתקדמות יותר עשויות לא ליפול בגדרי החריגים הסטטוטוריים, והסדר ספציפי עלול שלא במתכוון לייצר תמריץ שלילי לאימוץ של טכנולוגיות חדשניות יותר שיתפסו כמותרות מ"נמל המבטחים" הסטטוטורי. מטעם זה סבורנו כי רגולציה רכה של הסוגיה באמצעות חוות דעת ולא שינוי חקיקה היא מתווה נכון יותר לפתרון סוגיה זו. הפרק הבא יעסוק בהיקפה של חוות דעת זו ובסייגים לתחולתה.

ה. היקפה של חוות הדעת והסייגים לה

כאמור בפרק הקודם, הפרשנות המוצעת בחוות דעת זו היא כי האיזונים הקבועים בחקיקה הקיימת מגנים על פעולות הכרוכות ביצירת מאגר נתונים מתוייג לצורך למידת מכונה. על בסיס פרשנות זו, רשאים מיזמים של בינה מלאכותית לעשות שימוש בתכנים מוגנים על מנת לייצר מאגר נתונים לצורך אימון מערכות של בינה מלאכותית, בהתאם לאמות המידה המפורטות לעיל, כאשר יתכן כי במקרים מסויימים (כמו למשל שימוש ביצירות של יוצר אחד לשם הפקת פלט שמחקה את היצירות של אותו יוצר), תהליך למידת המכונה לא יזכה להגנה מכח חוות הדעת. פרק זה יבהיר את היקפה של חוות דעת זו ואת גבולותיה. להלן יידונו בקצרה השימושים המכוסים בחוות דעת זו ואלה המותרים ממנה.

1. שיפור וניטור של מערכות בינה מלאכותית

מיזמי בינה מלאכותית נדרשים למאגר הנתונים שיצרו גם לאחר שלב האימון הראשוני של המערכת ולאחר שהמערכת כבר פועלת, לצרכי ניטור ושיפור המערכת. אכן, הגם שמערכות בינה מלאכותית הן אוטונומיות בביצוע תהליך הלמידה והמשימות הנובעות ממנו,¹³⁰ אין מדובר במערכות "שגר ושכח". ללא ניטור, לימוד מחדש ושיפור ביצועים מתמיד, המערכת עשויה להתפתח בכיוונים לא צפויים ולהניב תוצאות אנומליות.¹³¹

מובהר אפוא בזאת כי חוות דעת זו חלה גם על שימוש מאוחר במאגר הנתונים לצרכי ניטור ושיפור המערכת כמפורט לעיל. כך, שימוש במאגר הנתונים או בחלק ממנו לצורך ניטור של ביצועי המערכת חוסה גם הוא תחת הסדרי השימושים המותרים, ככל שהשימוש המקורי חוסה תחת הגנות אלה. למעלה מן הצורך יוער כי במקרה זה, ברי כי אין תחולה להגנה בדבר שימוש זמני, מכיוון שמאגר הנתונים נשמר במערכת ולא נמחק בתום השימוש, ובשל כך לא יוכל להחשב זמני.

2. שימוש במאגר הנתונים או בנגזרות שלו על מנת לאמן מערכת אחרת

האם מאגר מידע שנוצר על בסיס חוות דעת זו ניתן לשימוש רק עבור המיזם שיצר אותו או גם עבור מיזמים אחרים? מחד גיסא, העברה של מאגר נתונים מתוייג בין מיזמים יכולה להוות פרקטיקה יעילה. יצירה של מאגר נתונים מתוייג מהווה כאמור תהליך עתיר משאבים, ועל מנת להשיא את התועלת הצפויה מפרוייקטים של בינה מלאכותית ולמנוע בזבוז, יש טעם ליתר את הצורך לייצר מחדש מאגר נתונים שכבר קיים. הרי בשכפול של המאמץ ליצור מאגר נתונים אין כל תועלת לבעלי זכויות היוצרים (שממילא על פי חוות דעת זו לא יפיקו רווח מכלילת יצירת במאגר נתונים נוסף). מאידך גיסא, מתן אפשרות להעביר מאגרי נתונים בין חברות ללא הגבלה עלול לתמרץ יצירה של פרקטיקות עסקיות שיסיטו את נקודת האיזון ויביאו לפגיעה בלתי מאוזנת בבעלי זכויות יוצרים. לאור ניתוח תכליתי זה, ראוי בעינינו להבחין בין מספר מצבים.

א. שימוש במאגר על ידי אותו מיזם או מתן שירות למיזם (לרבות מיקור חוץ). מצב אחד החוסה לדעתנו בבירור תחת האמור בחוות דעת זו הוא כאשר הגורם שיצר את המאגר מבקש לעשות בו שימוש נוסף בעצמו כדי לאמן מערכת למשימה אחרת, קשורה או לא קשורה למערכת המקורית. שימוש זה לדעתנו לא יהווה הפרה של זכויות יוצרים באם השימוש המקורי היה שימוש

¹³⁰ ראו לעיל פרק ב.

¹³¹ ראו למשל מערכת הבינה המלאכותית של גוגל, שזיהתה פנים כהי עור כפני גורילות. James Vincent, *Google 'fixed' its racist algorithm by removing gorillas from its image-labeling tech*, VERGE (Jan 12, 2018), <https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai>.

מדינת ישראל משרד המשפטים

מותר. אין טעם לחייב את המיזם ליצור שוב את אותו מאגר נתונים שכבר יצר. הדבר אינו יעיל (ואינו אכיף) וממילא אין בו כל תועלת עבור בעלי זכויות יוצרים, שלא יקבלו תשלום גם עבור יצירת המאגר בשנית. מצב נוסף שאינו מעורר בעיה הוא העברה של מאגר הנתונים לחברת מיקור חוץ כדי שתבצע שירותים עבור המיזם, כמו ארגון מאגר המידע, תיוג, אימון מחדש של המערכת, ניטור אלגוריתמי שלה וכיוצא באלה.

ב. פרויקטי תשתית. בפרוייקטים מסויימים של למידת מכונה, המשימה של המערכת הינה יצירה של מודל מאומן שנועד להוות תשתית לפיתוח של מערכות בינה מלאכותית אחרות. דוגמה מאלפת לפרוייקט כזה הינו תכנית ה-NLP הלאומית. תכנית זו נהגתה משזיהתה המדינה כשל שוק בתחום של עיבוד שפה טבעית בעברית ובערבית, הנובע בין היתר ממיעוט המשתמשים בשפות אלה ביחס לשפות אחרות. כשל זה מביא לפגיעה בכמות ובאיכות של מערכות מבוססות NLP בעברית ובערבית ביחס למקבילותיהן באנגלית. אחת המטרות הראשיות של התכנית הממשלתית היא לפתח מודל מאומן שעל בסיסו יוכלו מיזמים אקדמים, ממשלתיים ופרטיים לפתח שימושי NLP ספציפיים בעברית ובערבית.¹³²

פרוייקט תשתיתי כגון דא יוכל להשיג את מטרתו באופן מיטבי אם גם מערכות שיתבססו על המודל המאומן שיווצר במסגרתו יוכלו לעשות שימוש במאגר הנתונים שיצר או בנגזרות שלו. תחת פרשנות אחרת, יאלצו כל אחת מהמערכות הנשענות על המודל המאומן לנתב משאבים לפיתוח של מאגר נתונים משלהן (או לפתח את אותו מאגר הנתונים מחדש) במקום לפיתוח המסויימים של המערכת שלהם, ויהיו בכך משום בזבוז משאבים, חוסר יעילות ואף פגיעה באיכות המוצר הסופי. מאידך, הפגיעה בבעלי זכויות יוצרים כתוצאה משימוש נגזר במאגר המידע היא מצומצמת, אם בכלל קיימת. לאור זאת, חוות דעת זו מחילה את הסדרי השימושים המותרים גם על שימוש במאגר המידע על ידי חברות אחרות בפרוייקטים נגזרים.

ג. העברת מאגר הנתונים בין מיזמים בלתי קשורים. מקרה מורכב יותר להכרעה הוא כאשר אין מדובר בפרוייקטים תשתיתיים, במתן שירות או בפרוייקטים של אותו מיזם. ישנם שני מצבים אפשריים. האחד הוא העברה של המאגר בין חברות בינה מלאכותית בלתי קשורות. מקרה זה פחות מדאיג בעינינו שכן שיקולי תחרות ושוק מגבילים ממילא פרקטיקה כזו (וככל שהיא תינקט, צריך יהיה לבחון אותה לפי מבחני השימוש ההוגן כפי שהדבר נעשה כיום).

המצב השני הוא שחוות דעת זו תפורש בטעות כמתירה מודל עסקי שכל מהותו היא יצירת מאגרי מידע המורכבים מיצירות מוגנות ומכירתם לצדדים שלישיים. ולא היא. לא הרי שימוש במאגר הנתונים לצורך אימון מכונה כהרי שימוש שהוא בעיקרו עסקה במאגר עצמו. חוות הדעת לא נועדה לאפשר ליוצרי המאגר לעשות שימוש בדיני זכויות יוצרים כדי להפיק רווח ישיר או עקיף ממכירת המאגר עצמו (או מתן רשיון בו), בעודם נשענים על דוקטרינת השימוש ההוגן על מנת לשלול רווח בגין העסקה מבעלי זכויות היוצרים ביצירות המרכיבות את המאגר.¹³³

לאור האמור, מובהר בזאת כי הפרשנות המוצעת בחוות דעת זו להסדרי השימושים המותרים ביצירות מוגנות מאפשרת העברה של מאגר הנתונים בתמורה או שלא בתמורה לגופים קשורים, כגון בפרוייקט-תשתית או בעת מתן שירותים למערכות קיימות. אולם אין בחוות הדעת להכשיר מיניה וביה כל העברה של מאגר נתונים בין צדדים לא קשורים. במצבים כאלה, יבחנו בתי המשפט אד הוק את אופי השימוש, את שאלת התמורה ואינדיקציות נוספות להוגנות השימוש. מעל לצורך יוער כי עסקה במאגר הנתונים עצמו לא תוכל להחשב לשימוש אגבי ביצירות המרכיבות אותו. כמו כן, בדומה לתת הפרק הקודם, גם כאן לא יחול החריג בדבר העתקה זמנית, מכיוון שמאגר הנתונים לא יוכל להחשב זמני אם הוא לא נמחק בתום השימוש אלא מועתק ומועבר לצד שלישי. מעל לצורך יובהר שוב כי על מאגרי נתונים יכולים לחול דינים שונים לבר מדיני זכויות יוצרים, אך חוות הדעת מתייחסת אך להיבטים הנוגעים לזכויות יוצרים.

3. העמדה לרשות הציבור של מאגר הנתונים או מסירתו לצדדים שלישיים

¹³² ראו ממשלת ישראל, הודעה לעיתונות - הקמת איגוד חברות לטכנולוגיות שפת אנוש (NLP) בעברית ובערבית, <https://www.gov.il/he/departments/news/digital-nlp> (22.09.2020).

¹³³ מאגר מידע מקורי עשוי להיות מוגן כיצירה ספרותית מסוג לקט, כהגדרתו בס' 4 לחוק זכות יוצרים.

מדינת ישראל

משרד המשפטים

כפי שאין בחוות דעת זו כדי להכשיר לאור דיני זכויות יוצרים העברה של מאגר הנתונים לצדדים שלישיים ללא הגבלה, כך אין חוות דעת זו מכשירה אפריורית העמדה של מאגר המידע לרשות הציבור (גם באתרים ייעודיים לשימוש מכונה). הטעם לכך הוא כפול. ראשית, כאמור, חוות דעת זו לא נועדה להתיר מודל עסקי שמהותו היא ביצוע עסקאות במאגרי נתונים, והדבר יכול שיעשה גם באמצעות העמדתם לרשות הציבור. שנית, לגוף שהעמיד את היצירה לרשות הציבור לא יכולה להיות שליטה במעשיו של מקבל גישה למאגר הנתונים. יתכן שבסופו של דבר ייעשה בתכנים המרכיבים את מאגר הנתונים שימוש שאינו לצורך אימון מערכות לומדות. טול, לדוגמה, מיזם שאימן מערכות לייצר תקצירים של מאמרים אקדמיים ולשם כך, העתיק למאגר הנתונים מאמרים אקדמיים רבים. העמדת מאגר הנתונים לרשות הציבור, גם בפלטפורמות שנועדו למפתחי מערכות,¹³⁴ יאפשר למנועי חיפוש להפנות לפלטפורמות אלה משתמשים שמחפשים גישה למאמרים עצמם, ולפגוע בבעלי הזכויות במאמרים, אשר מציעים אותם בפלטפורמות משלהם.

ראוי להקדיש מספר מילים ליחס בין תת הפרק הזה לתת הפרק הקודם שעסק בהעברת מאגר הנתונים לצדדים שלישיים. שאלה חשובה העולה משילוב תתי-הפרקים הללו היא מהי מידת אחריותו של מיזם שהעביר את מאגר הנתונים לגוף שלישי לשם למידת מכונה באופן מותר, והגוף השלישי העמיד אותו לרשות הציבור או עשה בו שימוש שאינו לצורך למידת מכונה. דיון מעמיק בשאלה זו חורג מגבולותיה של חוות דעת זו. בקצרה יצויין כי אין בחוות דעת זו כדי לסטות מדוקטרינות בדבר הפרה שניונית של זכויות יוצרים כפי שנקבעו בחוק ובפסיקה.¹³⁵

4. תוצרי למידת המכונה

למידת מכונה אינה אלא אמצעי לאימון מערכות לבצע משימות. בסופו של דבר, המערכת מניבה תוצר כלשהו - יהא זה ניבוי, זיהוי, סיווג, שיערוך או יצירה חדשה המבוססת על למידת דפוסים מהיצירות שהוזנו למערכת בשלב הלמידה. תת-פרק זה מפנה את הזרקור לשימוש האחרון: מערכת שיוצרת יצירות בעצמה. דוגמאות ישנן למכביר. מערכת בינה מלאכותית אחת מייצרת שירים בסגנון נתן אלתרמן,¹³⁶ אחרות מחוללות ציורים "של" ואן גוך¹³⁷ או "של" רמברנדט,¹³⁸ מערכת נוספת כותבת מאמרים אקדמיים¹³⁹ ועוד אחת מבצעת שירים של יוצר אחד בסגנונו של יוצר אחר.¹⁴⁰

יצירות שנוצרו על ידי מכונה מעוררות שאלות בדיני זכויות יוצרים החורגות מגבולותיה של חוות דעת זו, כגון בעלות בתוצרי המכונה.¹⁴¹ לצורך חוות דעת זו תובהר נקודה אחת בלבד: ההגנות שנדונו לעיל באשר ליצירת מאגרי נתונים אינן חלות באופן אוטומטי גם על התוצר של מערכות מבוססות בינה מלאכותית. במילים אחרות, **יצירה לא תחשב כיצירה שאינה מפרה זכויות יוצרים רק מן הטעם שהיא נוצרה על ידי מערכת מבוססת בינה-מלאכותית.** את בחינת חוקיות התוצר יש לעשות כדרך של יצירות אחרות. כך, יש לבחון אם תוצר למידת המכונה מפר את אחת הזכויות האקסקלוסיביות המנויות בסעיף 11 לחוק זכות יוצרים. אם התשובה היא חיובית, יש לבחון את הסדרי השימושים המותרים אד הוק, ביחס לשימוש הספציפי של מערכת הבינה המלאכותית ביצירה.¹⁴²

¹³⁴ ראו למשל [Github](https://github.com), <https://github.com>.

¹³⁵ ראו סעיפים 47, 48, 48א, 51 לחוק זכות יוצרים; ע"א 5977/07 האוניברסיטה העברית בירושלים נ' בית שוקן להוצאת ספרים בע"מ (נבו) 20.6.2011.

¹³⁶ ראו המחוללתרמן - מחולל שירי אלתרמן אוטומטי, <http://altermanator.herokuapp.com>.

¹³⁷ ראו https://ouchhh.tv/ai-van-gogh_immersive-data-painting#:~:text=STATEMENT%20%3E-All%20of%20Van%20Gogh's%20works%20created%20during%20his%20lifetime%20were,works%20were%20brought%20to%20life.

¹³⁸ ראו למשל [The Next Rembrandt](https://www.nextrembrandt.com), <https://www.nextrembrandt.com>. לדוגמה מרכזית נוספת ראו גם [Abraham Project](https://abraham.ai), <https://abraham.ai>.

¹³⁹ ראו [BoredHuman.com](https://boredhumans.com/research_papers.php), https://boredhumans.com/research_papers.php.

¹⁴⁰ ראו Knight, לעיל ה"ש 28.

¹⁴¹ ראו לעיל ה"ש 37, 16.

¹⁴² על אף ההבחנה האנליטית בין הקלט של המערכת לפלט שלה, מבחינה מעשית אין לייצר "חומות סיניות" בין השלבים, וזאת כדי לאפשר תביעה כנגד מיזם בגין תוצר של מערכת הבינה המלאכותית שלו. "חומות סיניות" נוקשות יאפשרו למיזם לחמוק מאחריות על בסיס הטענה כי שלב הלמידה מהווה שימוש מותר, ואילו שלב הפלט נוצר על ידי המערכת ללא מעורבות של המיזם. המערכת עצמה כמובן אינה אישיות משפטית ואינה יכולה להיבע. על מנת למנוע זאת, לצורך ביסוס אחריות בזכויות יוצרים, יש לאפשר בחינה של פעילות המיזם כיחידה אחת במקרים מתאימים, מבלי להבחין בין שלב האימון בו היה מעורב גורם אנושי לשלב יצירת התוצר שנעשה על ידי המחשב באופן אוטונומי. ראו גם לעיל ה"ש 37; 537, 593 (2021) 58 Hous. L. Rev. Nadia Banteka, *Artificially Intelligent Persons*.

מדינת ישראל

משרד המשפטים

5. תניות חוזיות ואמצעי הגנה טכנולוגיים

החלת הסדרי השימושים המותרים על למידת מכונה נועדה לקבוע את מערכת האיזונים הראויה בין זכויות יוצרים לזכויות משתמשים בתחום זה. תת-פרק זה מתמודד עם החשש שמטרה זו תושם לאל באמצעות הטלת הגבלות חוזיות או טכנולוגיות על איסוף מידע לטובת מאגרי נתונים.¹⁴³ כך, למשל, בעלי זכויות היוצרים או פלטפורמות אינטרנטיות המנגישות את המידע לציבור עלולים לסכל את הפרשנות דכאן באמצעות תנאי שימוש האוסרים איסוף מידע או להגן על המידע באמצעים טכנולוגיים שיסכלו העתקה. במצב דברים זה, יהא ההיתר המשפטי דכאן תאורטי בלבד.¹⁴⁴

השאלה אם ניתן להתנות בחוזה על דיני זכויות היוצרים ובפרט על ההגנות שהם מספקים למשתמשים נדונה לא מעט ברחבי העולם בהקשרים מגוונים.¹⁴⁵ בהקשר הנדון, החשש מפני סיכול כאמור של איסוף מידע הביא את הפרלמנט האירופי שדן בסוגיה למסקנה כי יש לקבוע בחקיקה שהזכות לאיסוף מידע לצורך בניה של מאגרי מידע ללמידת מכונה תגבר על טענות חוזיות של בעלי זכויות או של פלטפורמות המחזיקות במידע.¹⁴⁶

בישראל אין פסיקה מפורשת בשאלה אם ניתן להתנות בחוזה על הסדר השימוש ההוגן. נקודת המוצא לדיון היא עקרון חופש החוזים.¹⁴⁷ ברם, למרות מרכזיותו של עקרון חופש החוזים, כוחם של הצדדים להתנות על הסדרים המגשימים ערכים חברתיים החורגים מעניינם הפרטי הוא מוגבל.¹⁴⁸ התנאה על הסדר מעין זה כפופה לדינים העשויים להגבילה, ובפרט לדוקטרינות החוזיות של תום לב ותקנת הציבור, שעשויות לשלול תוקפן של הסכמות הסותרות אינטרסים ציבוריים וחברתיים חיוניים.¹⁴⁹ ולמנגנוני פיקוח מכוח חוק החוזים האחידים, התשמ"ג-1982.¹⁵⁰

נדון תחילה בהגבלות חוזיות על איסוף מידע שמקורן בחוזה אחיד, כגון תנאי שימוש באתר הנקבעים באופן חד צדדי. חוות דעת של מחלקת ייעוץ וחקיקה (אזרחי) עסקה בסוגיה של התנאה על שימוש הוגן בחוזה אחיד בשנת 2016.¹⁵¹ נקודת המוצא בחוות הדעת האמורה כמו גם בבתי המשפט היא כי האופי ה-"take it or leave it" של תנאי השימוש באתר עושה אותו ברגיל לחוזה

¹⁴³ ראו Christophe Geiger, *The Answer to the Machine should not be the Machine*, EIPR(4) 121 (2008); ראו גם אלקין-קורן, לעיל ה"ש 72.

¹⁴⁴ Contractual Lucie Guibault (2002), Copyright Limitations and Contracts: An Analysis of the Overridability of Limitations on Copyright, The Hague, Netherlands: Kluwer Law International. עבור המיזם, תביעה חוזית בכל זאת תהא עדיפה על פני חשיפה בזכויות יוצרים שבצידה סעד של פיצוי ללא הוכחת נזק. ראו לעיל ה"ש 47.

¹⁴⁵ ראו גם אלקין קורן, לעיל ה"ש 49. דיון נרחב התעורר בארה"ב בעקבות פסק הדין בעניין ProCD, Inc. v. Zeidenberg, 86 F.3d 1447 (7th Cir. 1996). שאלה זו משיקה לשאלה שחורגת מתחומה של חוות דעת זו, האם החריגים לחוק זכות יוצרים מהווים זכות או טענת הגנה בלבד. בקנדה הוכרה הדוקטרינה של זכויות משתמשים ב-Canadian Ltd. v. Law Soc'y of Upper Canada, 2004 SCC 13, [2004] 1 S.C.R. 339 (Can.). מאוחרים יותר. בישראל נדחתה תפיסה זו בע"א Premier League, לעיל ה"ש 21 (פסקה 17-18) ("...יש הסבורים כי ניתן לראות בשימושים המותרים על-פי החוק החדש, כזכויות של המשתמשים, במובן זה שהן עשויות לשמש גם כטענה פוזיטיבית, ולא רק כטענת הגנה. אלא שאין בידי לקבל טענה זו."). ראו גם JOYCE C., PATRY W., LEAFFER M. & JASZI P., COPYRIGHT LAW (4th ed.) 715, 1998 (מתייחסים לשימוש הוגן באותו עמוד כ-"privilege", "limitation" ו-"affirmative defense").

¹⁴⁶ ראו European Parliament, לעיל ה"ש 46, בעמ' 12 ("In order to promote TDM research, the EU legislator should clarify that protection against contractual and technological override should also be always extended to TDM mining both materials protected and not protected by IPRs, including those made available in a database."). הבהרה כזו נדרשת באיחוד האירופי כיוון שבאירופה נפסק שמותר לקבוע תנאי שימוש באתר גם מעבר להגנת זכויות יוצרים. CJEU, C-30/12, Ryanair Ltd v. PR Aviation BN (15 January 2015). ECLI:EU:C:2015:10.

¹⁴⁷ ראו גבריאלה שלו ואפי צמח דיני חוזים (מהדורה רביעית 2019), פרק 2. ראו גם אלקין-קורן, לעיל ה"ש 49, בעמ' 374-375; תמיר אפורי, חוק זכות יוצרים, עמ' 193-194 (2012).

¹⁴⁸ ראו שלו וצמח, שם, שם.

¹⁴⁹ ראו שלו וצמח, שם, בעמ' 526 (התנאה על ההסדר בדבר פירוש חוזה); גבריאלה שלו ויהודה אדר דיני חוזים – התרופות (2009) בעמ' 91 ואילך (התנאה על דיני תרופות בשל הפרת חוזה). לתקנת הציבור באופן כללי ראו שלו וצמח, שם, בעמ' 640 ואילך.

¹⁵⁰ ראו חוק החוזים האחידים, תשמ"ג-1982. ראו גם ורדה לוטסהוויז וטנה שפניץ, חוזים אחידים, עמ' 42 (התשנ"ד-1994).

¹⁵¹ ראו משרד המשפטים, מחלקת ייעוץ וחקיקה (אזרחי), היחס בין שימושים מותרים לפי חוק זכות יוצרים, לבין תניות בחוזה אחיד האוסרות שימושים אלה, כ"ט סיוון תשע"ו. ניתן לצפיה ב-https://www.gov.il/BlobFolder/generalpage/legal-opinions-01/he/legal-opinions_lo-cl-18-7-2016.pdf.

מדינת ישראל משרד המשפטים

אחיד.¹⁵² משמעות הדבר היא שתניות מקפחות בתנאי שימוש באתר לא יוחזקו כתקפות.¹⁵³ ישנו בסיס איתן לסברה כי תניה האוסרת על איסוף מידע באופן קטגורי תוחזק כתניה מקפחת, בין היתר כיוון שהיא מנוגדת להוראת החוק בדבר שימוש הוגן, המאפשרת לפחות עקרונית איסוף תכנים אוטונומי שיחשב לשימוש מותר בתכנים.¹⁵⁴

הסוגיה מורכבת יותר במקרים בהם החוזה אינו אחיד אלא נקבע באמצעות משא ומתן בין הצדדים. מחד, במקרים אלה, קל יותר להעלות על הדעת מקרים בהם ויתור מצד המשתמש על השימוש בהסדרי השימושים המותרים יהיה מוצדק.¹⁵⁵ מאידך, במקרים רבים, ויתור כאמור עלול לייצר החצנות שליליות או כשלי שוק אחרים. כאמור, השאלה אם הסדר השימוש ההוגן הוא קוגנטי או בר-התנאה היא רחבה והכרעה בה חורגת מגבולותיה חוות דעת זו. על דרך ההכללה, הסדר יפורש כקוגנטי כאשר הוא נועד לקדם אינטרס ציבורי שיסודו בטובת הכלל והוא חורג מהאינטרס האישי של המתקשרים בחוזה הקונקרטי. העקרון העומד ביסוד הכלל הוא שכיוון שההסדר אינו כפוף לרצונם של מתקשרים חוזיים קונקרטיים, אין לתת בידם לסכל את המטרה שביסוד ההסדר.¹⁵⁶ בעניין דנן, אם יוכרע בעתיד כי הסדר השימוש ההוגן הוא קוגנטי, התנאה עליו תעמוד, לכאורה, בניגוד לתקנת הציבור ועשויה להיות בטלה מכוח הוראות סעיף 30 לחוק החוזים הכללי.¹⁵⁷ כאמור, אין אנו מבקשים לקבוע מסמרות בעניין במסגרת זו. ברם, גם ללא הכרעה בדבר הקוגנטיות של הסדר השימוש ההוגן, בידי בתי משפט כלים "רכים" יותר לפרשנות חוזית תכליתית, שבכוחם לאפשר יצירה של מאגרי מידע ולמנוע התנאה על האיזונים הרצויים בסוגיה המשפטית-טכנולוגית הנדונה, במקרים המתאימים.¹⁵⁸

ה. אחרית דבר

מטרתו של מסמך זה היא לשרטט את גדרי השימושים המותרים בתהליך למידת המכונה. מסקנתו היא כי יצירה של מאגר נתונים המורכב מיצירות מוגנות לצורך אימון מכונה תחסה לרוב תחת דוקטרינת השימוש ההוגן, ולעתים גם תחת דוקטרינות השימוש האגבי או ההעתקה הזמנית, כמפורט לעיל. על פי הניתוח המוצע במסמך זה, פרשנות שתשלול את ההגנות האמורות תסכן את היתרונות הגלומים בלמידת מכונה ואת מעמדה של ישראל כמובילה עולמית בתחום, וזאת מבלי לייצר יתרונות של ממש לבעלי זכויות יוצרים. גם העובדה כי מדינות העולם שנתנו לעניין זה את דעתן הכריעו רובן ככולן כי על דיני זכויות היוצרים לאפשר יצירת מאגרי מידע ללמידת מכונה מעידה על מלאכת האיזונים המקובלת בעולם בהקשר זה ומחזקת את עמדתנו כי חוות דעת זו בצדק יסודה. מאידך, מצבים מיוחדים - שלא כולם ברי צפיה בעת הנוכחית - יצדיקו סטייה מן הכלל האמור ויסיטו את נקודת האיזון לעבר הגנה יתרה על זכות היוצרים. כמפורט לעיל, מסמך זה אינו מבקש להסדיר את מעמדו החוקי של **תוצר** למידת המכונה, מתוך הבנה שקיימים - ועוד יתקיימו - סוגים מגוונים של תוצרים כאמור, ולא מן התבונה לקבוע מסמרות בעניין זה. אשר על כן, מעמדו החוקי של תוצר שנוצר מלמידת מכונה לא יהנה מנמל מבטחים קטגורי ויבחן אד-הוק במסגרת הנוהגת.

ישאל השואל אם נדרשים תיקוני חקיקה או התקנת תקנות - למשל לפי סעיף 19(ג) לחוק - על מנת לקבע את הפרשנות המוצעת בחוות דעת זו ולהגביר ודאות בשוק הבינה המלאכותית. מספר מדינות, כאמור, בחרו בדרך זו להסדרת הסוגיה. עם זאת, בשלב זה הכרענו כי מוטב שלא להידרש לתיקוני חקיקה. אמנם חקיקה תייצר ללא ספק רמה גבוהה יותר של ודאות עבור יוצרים, מיזמי בינה מלאכותית ושחקנים נוספים. אך אליה וקוץ בה. עד להשלמת הליכי החקיקה, הן הטכנולוגיה והן תנאי השוק צפויים להשתנות דרמטית, וממילא יתעורר שוב הצורך בכלים פרשניים. הדבר נכון גם באשר להתקנת תקנות, גם אם בעוצמה פחותה. אכן, ישנו קושי אינהרנטי בקידום חקיקה שהיא טכנולוגית-ספציפית בזירה דינאמית כמו תעשיית למידת המכונה. בזירה זו, הסדרים ספציפיים

¹⁵² ראו רע"א 5860/16 Facebook Inc נ. אוהד בן חמו (31.05.2018).

¹⁵³ ראו ס' 19 לחוק החוזים האחידים תשמ"ג-1982.

¹⁵⁴ ראו לעיל, ה"ש 152. ראו גם לוסטהוויז ופנניץ, לעיל, ה"ש 150, בעמ' 83 (מורות כי סטייה מהוראות חוק - גם אם דיספוזיטיביות - נחשבת כבלתי סבירה לכאורה, ותוחזק כתנאי מקפח). Katherine J. Strandburg, *Free Fall: The Online Market's Consumer Preference Disconnect*, 2013 U. CHI. LEGAL F. 95 (2013); Lital Helman, *Pay for Privacy (Privacy) Performance: Holding Social Network Executives Accountable for Breaches in Data Privacy Protection*, 84 BROOK. L. REV. 532, 537-38 (2019).

¹⁵⁵ ראו למשל גרינמן, לעיל ה"ש 12, בעמ' 333-334.

¹⁵⁶ ראו, למשל, ע"א 156/82 ליפקין נ' דור הזהב פ"ד לט" (3), 85, 94 (1985); ד"נ 22/82 בית יולס בע"מ נ' רביב משה פ"ד מג(1) 441, 463 (1989).

¹⁵⁷ ראו, למשל, ליפקין, שם, בעמ' 95. השוו: ע"א 11/84 רבינוביץ' נ' שלב, פ"ד מ(4), 533, בעמ' 547 (1986).

¹⁵⁸ ראו לעיל ה"ש 147 וטקסט נלווה.

מדינת ישראל משרד המשפטים

למצב טכנולוגי ושוקי נתון עלולים בסבירות גבוהה להראות מיושנים במבט לאחור, לעתים זמן קצר מאד לאחר שנכנסו לתקפם.¹⁵⁹ הליכי החקיקה באירופה מהווים דוגמה בולטת לעניין זה. טרם הושלמו הליכי ההטמעה של הדירקטיבה החדשה, וכבר מסתבר שהדירקטיבה חסרה התייחסות להיבטים מרכזיים בתחום של למידת מכונה.¹⁶⁰

יתר על כן, בהקשר דנן, לחקיקה ספציפית עשוי להיות אפקט גרסיבי מבחינה טכנולוגית. דבר חקיקה שיעניק "נמל מבטחים" סטטוטורי לטכנולוגיות למידת מכונה הידועות למחוקק היום ייצר כמובן תמריץ חזק בשוק לעשות שימוש בטכנולוגיות המוגנות בחקיקה, על מנת להמנע מאחריות. תמריץ זה לא יאבד מכוחו כמלוא הנימה כשיחדרו לשוק טכנולוגיות חדשות. אך ככל שהטכנולוגיות החדשות יהיו מהפכניות יותר, כך יגבר הסיכון שהן לא יחסו תחת נמל המבטחים הסטטוטורי. לדבר עשוי להיות אפקט דינאמי שלילי, כאשר מפתחים של טכנולוגיות למידת מכונה לא יתמרצו לפתח (או למצער לפתח בישראל) טכנולוגיות חדשניות שעלולות לצאת מגדרי החוק, מחשש להתבע או מחשש שטכנולוגיות אלה לא יאומצו בשוק. כך, בעוד במגוון הקשרים טכנולוגיים, מערכת האיזונים מצדיקה ולפעמים אף מחייבת התערבות חקיקתית, בהקשר דנן, אנו רואים את תפקידנו בהתוויית מערכת האיזונים בצורה דינאמית שתייצר תמריצים נכונים עבור השחקנים השונים כדי להשיא את התועלת החברתית.

סיכומו של דבר, בניגוד לשיטות משפט אחרות, דיני זכויות היוצרים בישראל מאפשרים פיתוח משפטי דינאמי ללא שינויי חקיקה, באמצעות דוקטרינות דוגמת השימוש ההוגן. ליכולת זו יש ערך שלא יסולא בפז בהקשר של רגולציה בעולם טכנולוגי מתפתח. חוק זכות יוצרים כולל בו הגנות שיש בכוחן לייצר וודאות משפטית ברמה סבירה מבלי לקבע מבחנים סטטוטוריים שספק אם יעמדו במבחן הזמן. האפשרות לקדם חקיקה או להתקין תקנות נותרת בגדר אפשרות למקרה שלא יהיה די בחוות דעת זו ובפסיקה שלאחריה כדי לפזר את ערפל אי הוודאות האופף סוגיה זו. לעת עתה, עמדתנו היא כאמור שגם כיום למידת מכונה עומדת בתנאי החריגים לדיני זכויות היוצרים. אנו תקווה כי עמדתנו זו תתקבל גם בבתי המשפט ככל שהדברים יבואו לפתחו.

¹⁵⁹ ראו Lital Helman, *Curated Innovation*, 49 AKRON L. REV. 695 (2016).

¹⁶⁰ ראו לעיל פרק ד.