



TECHNION

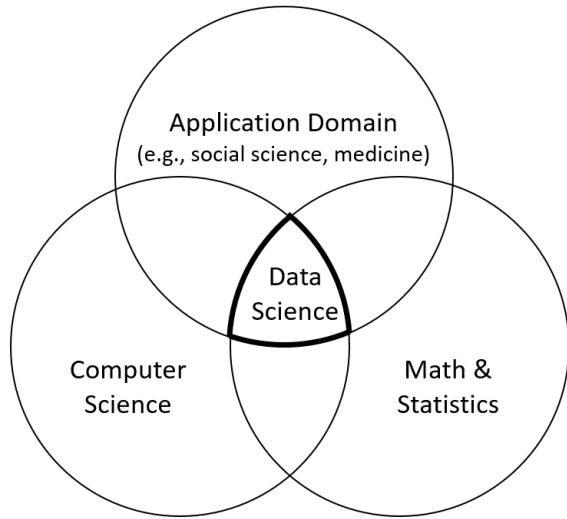
Israel Institute of Technology

Data (Science) Thinking

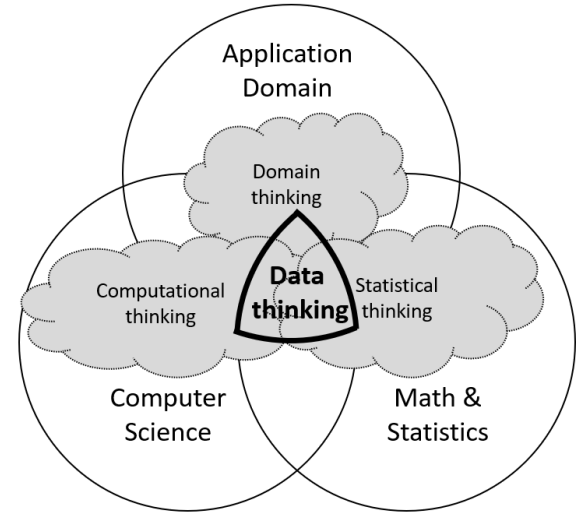
Methods of Teaching Data Science Course

Spring 2022

Data science and data thinking



(a) Data science integrates computer science, mathematics, and statistics, and a real-world domain.



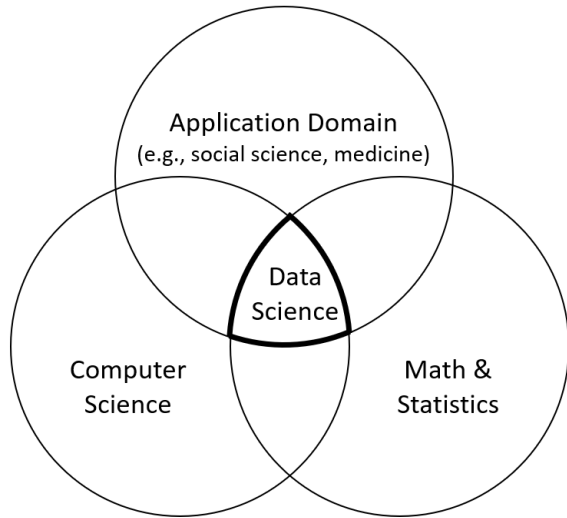
(b) Data thinking integrates computational thinking, statistical thinking, and domain thinking.

Data (Science) Thinking

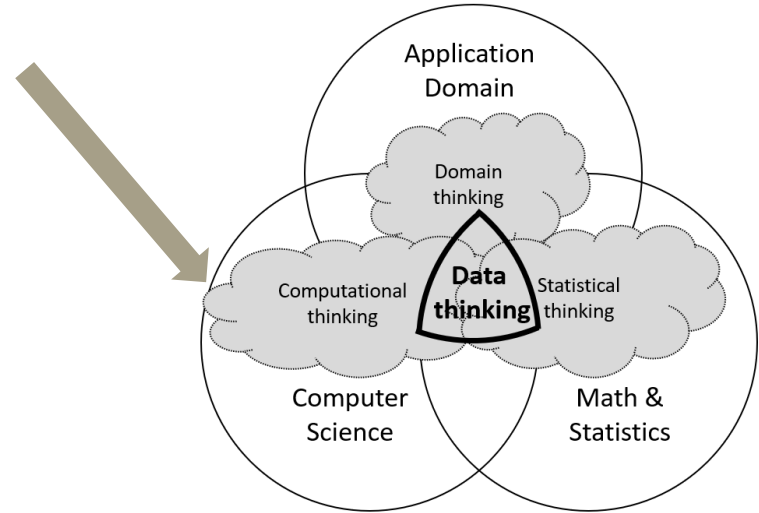
Data thinking: The mode of thinking required from data scientists (not only professional ones) that integrates the thinking modes associated with the disciplines that make up data science.

Discipline	Discipline Thinking
Computer science	computational thinking
Statistics	statistical thinking
Mathematics	object-process duality
Each application domain	thinking skills core principles ethical considerations

Data science and data thinking



(a) Data science integrates computer science, mathematics, and statistics, and a real-world domain.



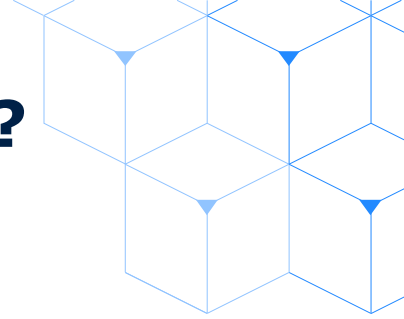
(b) Data thinking integrates computational thinking, statistical thinking, and domain thinking.

Computational thinking: What it is?



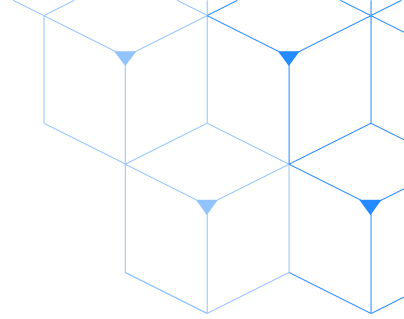
- ▶ Computational Thinking – Important readings
 - Papert, S. (1990). *Mindstorms: Children, computers and powerful ideas*. New York: Basic Books.
 - Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. <https://doi.org/10.1145/1118178.1118215>
- ▶ Computational thinking
 - is recognized today as one of the central **21st century skills**;
 - is useful and can **be applied in all disciplines**;
 - emphasizes **computer science problem solving skills** which promote learning experiences and support learning processes.
 - gives students **socio-economic benefits** and may close social gaps and **promote social mobility**.

Computational thinking: What does it include?



- ▶ Computational thinking includes:
 - **Cognitive skills** (Cuny et al., 2010)
 - problem formulation; dividing a problem into sub-problems; organization and logical analysis of data; representation of data with models and simulations by using abstraction; suggestion and assessment of several solutions to a given problem; examination and implementation of the chosen solution; and the generalization and transfer of the solution to a range of problems.
 - **Social skills** that support learning processes
 - teamwork; time management and planning; scheduling tasks.

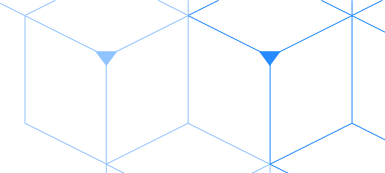
Computational thinking: How to teach it?



► Computational thinking is not:

- Computational thinking is about **thinking processes**, and therefore, **its implementation is independent of technology**.
- Teaching according to the computational thinking approach **does not emphasize the teaching of a specific subject matter**; rather, it emphasizes the acquisition of broad and multidisciplinary knowledge and set of skills that **can be applied in a variety of contexts**.
 - When **computational thinking** is integrated in all areas of studies, learners
 - **deepen their understanding of the discipline they study**
 - **develop their computational thinking skills (cognitive and social)**.
 - **This observation is important in the context of data science since data science itself is carried out in context – the application domain from which data is taken.**

CampusIL: Computational thinking MOOC



Orithazzan

קורסים אירועים תמיכה

המועצה לאשכול גבוהה
المجالس الأعلى مجلس
المستوى العالي للمجلس
מועצה לתכנון ולתקצוב

מערך הדיגיטל הלאומי
מטה ישראל דיגיטלית

חשיבה חישובית

פתרון בעיות מעולם לא היה קל יותר! הצטרפו לקורס שמלמד ומפתח כישורי חשיבה חישובית הניתנים ליישום בתהליכי פתרון בעיות בכל תחומי הדעת ותחומי החיים. בקורס נלמד להסתכל על תהליכים חישוביים, לנתחם ולפתח עבורם סימולציות בסביבת האנימציה סקראץ' (Scratch).

לעמוד הקורס

לומדים יקרים,

אנו שמחים שהצטרפתם לקורס "חשיבה חישובית". מטרת העל של הקורס היא פיתוח ידע מושגי וכישורי חשיבה חישובית המקובלים היום כנדרשים לכל אדם במאה ה-21. כישורי החשיבה החשובות מודגמים בקורס ביחס לתחומי דעת רבים ושונים. זאת, באופן דינמי, מונגש, תוך שימוש בסביבת הפיתוח סקראץ' (Scratch).

בקורס זה תלמדו:

- פיתוח גישות חשיבה, ביניהן: הגדרת בעיה, ניתוח בעיה, חלוקת בעיה לתת-בעיות, הפשטה, והכללה.
- פיתוח יכולת זיהוי וניתוח של תהליכים חישוביים הלקוחים מתחומי דעת שונים, תוך פתרון בעיות רלוונטיות לכל תחומי הדעת ופיתוח הסתכלות בינתחומית.
- פיתוח מיומנות טכנולוגית של בניית תסריטים בסביבת קוד מבוססת בלוקים.

הקורס מתאים לקהל הרחב בכל העילאים. עם וולא ניסיון קודם בפיתוח תוכניות מחשב. אתם מוזמנים להצטרף אלינו

CampusIL

Additional relevant courses

קורסים אירועים תמיכה

OrithHazzan

קמפוס IL

Self.py - הדרך שלך ללמוד פייתון

בקורס תרכשו בצורה פרקטית וחוויתית את כלי הבסיס לפיתוח קוד בשפת פייתון. שפת פייתון היא שפה אינטואיטיבית המאפשרת ביצוע פעולות מורכבות בפשטות, על כן שימושית מאוד בעולמות הסייבר.

[לעמוד הקורס](#)

קורסים אירועים תמיכה

OrithHazzan

קמפוס IL

מבוא למדעי המחשב בשפת C

בקורס זה תלמדו כיצד לפתח אלגוריתמים כדי לפתור בעיות מיחשוב באמצעות שפת התכנות C. בסיום הקורס תדעו לפתח תוכניות מחשב באמצעות שפה זו וגם תוכלו למדוד את יעילותם של האלגוריתמים והתוכניות אותם פיתחתם.

[לעמוד הקורס](#)

קורסים אירועים תמיכה

OrithHazzan

קמפוס IL

הכנה לביטוי מדעי המחשב C#

בחינת הבגרות במדעי המחשב מתקרבת - אל תגיעו לא מוכנים! הציגו ללמוד ולא יודעים מאיפה להתחיל? חסרו לכם תרגול? נתקעתם בנושא מסוים? הצטרפו אלינו לקורס הכנה לבגרות במדעי המחשב בשפת C# ותגיעו מוכנים לבחינה המתקרבת - הצטרפו עכשיו!

[לעמוד הקורס](#)

קורסים אירועים תמיכה

OrithHazzan

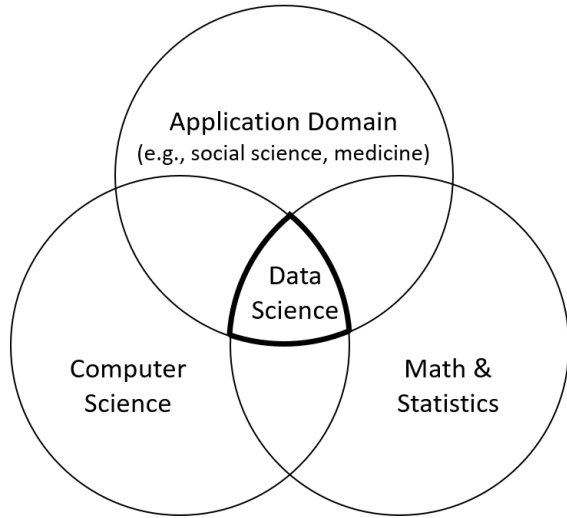
קמפוס IL

הפסיכומטרי של המדינה

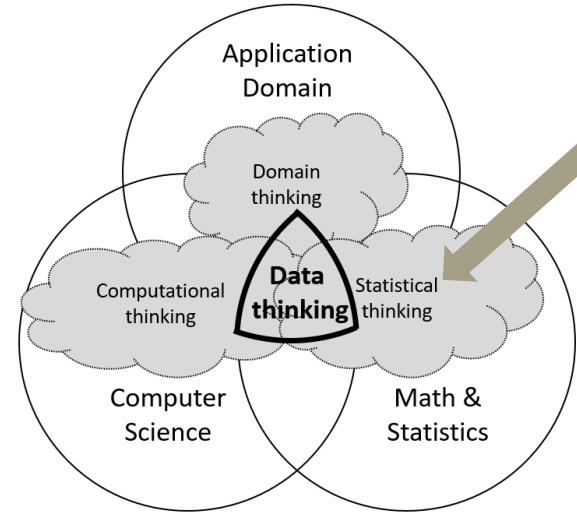
קורס הפסיכומטרי הרשמי של המדינה מטעם מערך הדיגיטל הלאומי, בשיתוף המרכז הארצי לבחינות ולהערכה (הגוף שמחבר ומעביר את הבחינה). קורס הכנה מלא, אינטראקטיבי, פתוח לכולם, ובחינם.

[לעמוד הקורס](#)

Data science and data thinking



(a) Data science integrates computer science, mathematics, and statistics, and a real-world domain.

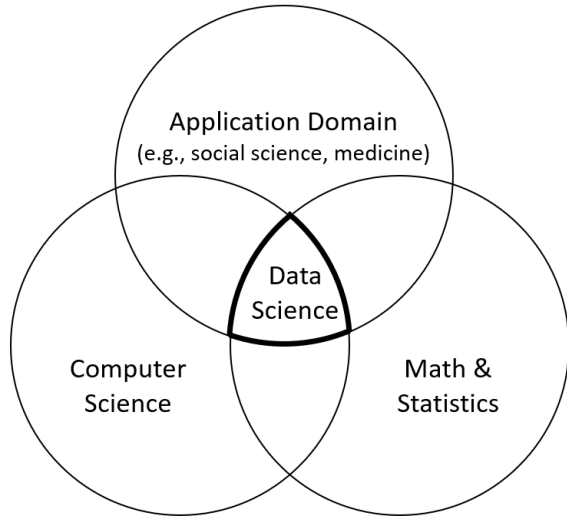


(b) Data thinking integrates computational thinking, statistical thinking, and domain thinking.

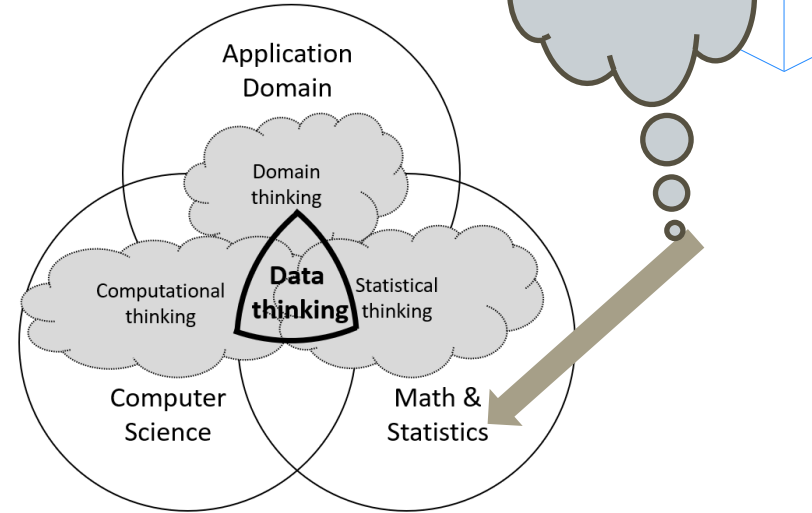
Statistical thinking

- ▶ Statistical thinking is associated with understanding **the essence, characteristics, and variability of real-life data**.
- ▶ Ben-Zvi & Garfield (2004): statistical thinking "*involves an understanding of why and how statistical investigations are conducted and the 'big ideas' that underlie statistical investigations*" (p. 8).
- ▶ Statistical thinking includes
 - (a) the understanding that **variation** exists in any data source and that real-life data contain outliers, errors, biases, and variance;
 - (c) the nature of **sampling** and how to infer from samples to populations;
 - (b) when and how to **use specific statistical data analysis methods**;
 - (d) statistical **models** and their usage;
 - (e) the **context** of a given problem when performing investigations and drawing conclusions;
 - (f) the **entire process** of statistical inquiry; and
 - (g) the relevance of **critique and evaluation** of inquiry results.

Data science and data thinking



(a) Data science integrates computer science, mathematics, and statistics, and a real-world domain.

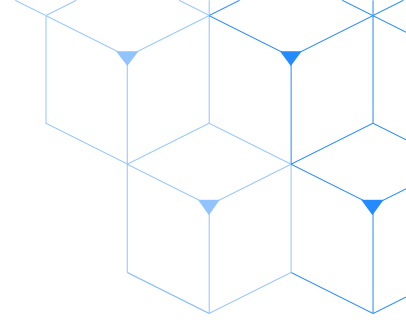


(b) Data thinking integrates computational thinking, statistical thinking, and domain thinking.

Mathematical thinking

- ▶ Write in the chat 3 mathematical concepts that you faced difficulties in their understanding.

Mathematical thinking



► The process-object duality theory

- Relevant for data science education due to the **centrality of the concepts of algorithm and data in data science.**
 - Wing (2006): “Computational thinking is [...] interpreting code as data and data as code.” (p. 33)
 - These cognitive theories has educational implications for **the design of pedagogical tools**
 - **Algorithms: their process and their properties**

Object-process duality: **The KNN algorithm**

- ▶ **How would you explain to a friend what the KNN algorithm is?**
- ▶ Discussion:
 - The question formulation
 - What can we learn from students' answers about their conception of the **KNN algorithm**?

Object-process duality: The KNN algorithm

Students' conception of the KNN algorithm as a process / object can be categorized by their examination of the following properties of the KNN algorithm

▶ Process conception:

- (P1) Calculate distance from all samples
- (P2) Pick the K nearest samples
- (P3) Find the label of the majority
- (P4) Tune the hyperparameter K to improve performance (to avoid underfitting and overfitting)

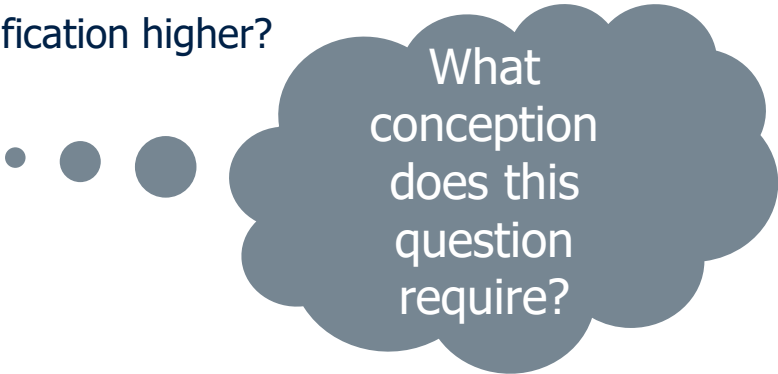
▶ Object conception: O1-O3 address the classification, O4 addresses the algorithm's performance and the O5-O7 address the algorithm's complexity (determined by the number of distance calculations):

- (O1) Classification depends on similarity
- (O2) Classification is determined by distance
- (O3) The classification of a specific unknown example depends on K
- (O4) The performance of the KNN algorithm for a specific K depends on the distribution of the data
- (O5) The number of distance calculations depends on the number of training samples
- (O6) The number of distance calculations depends on the number of features
- (O7) The number of distance calculations does not depend on K

Object-process duality: The KNN algorithm

In order **to classify dogs as Poodle or Labrador**, four characteristics were selected: height, weight, tail length, and ear length. The training set included 1,000 dogs, 500 of each kind. Based on this data set, we wish to classify an unknown dog using the KNN classifier.

- a. For $K=5$: How many times is the square operation executed?
- b. For $K=11$: How many times is the square operation executed?
- c. What conclusion can you draw from your answers to the above two questions?
- d. In your opinion, when are the chances of a correct classification higher?
 - I. $K=5$
 - II. $K=11$
 - III. It is impossible to decide
 - IV. I do not know
- e. Explain your answer.



What conception does this question require?

Mathematical thinking

► Object-process duality - The KNN algorithm

- חומרי לימוד קובי

<https://drive.google.com/drive/folders/1bfnoHo-95n1O1n8qORjWKI7iRzc27bv8>

בחרו את דף העבודה על KNN. זהו את העקרונות הפדגוגים שהנחו את פיתוחו.

Mathematical thinking: Reducing Abstraction



- ▶ Process-object duality is one perspective at the Reduction of the abstraction level phenomenon
- ▶ The reduction of the level of abstraction is illustrated by **the concept of data**:
 - Data with **specific features** are more concrete (less abstract) than **abstract features**
 - images with features “red” and “green” are more concrete than objects with features “ x_1 ” and “ x_2 ”.
 - Data with **specific values** are more concrete than **abstract values**
 - an image with features [107, 83, city] is more concrete than an object with features [x_1, x_2, x_3].
 - **Data** with meaning in the real world are more concrete than just **numbers**
 - “the level of red in the image is 107” is more concrete than the fact “the value of feature x_1 is 107”.
- ▶ **Note: Reducing the level of abstraction should be done very carefully.**
 - Reducing the level of abstraction too frequently in too many occasions may lead learners to conceive the specific (less abstract) case as the general (more abstract) case, a conception that in turn, may limit their problem-solving skills.

חומרי הלימוד של המורים (קובי)

מדעי המחשב אריאל בר-יצחק

CSITBlog

אתר הבלוגים של המורים

arib30@gmail.com

שלום אורח | התחבר | מנוי חדש | התנתק | לכל הבלוגים

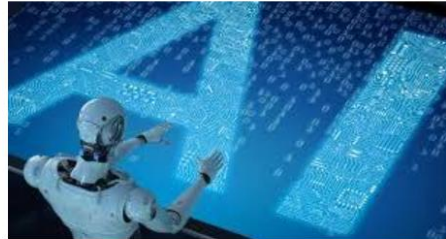
קהילות הבלוג

כרך:

מדעי המחשב

תוכן:

פתרונות בגרות מדעי המחשב
למידה עצמית לבגרות
מודלים חישוביים ותוכנת JFLAP
סביבת Visual Studio
למידת מכונה חומרי לימוד
למידת מכונה סרטונים חלק א
למידת מכונה סרטונים חלק ב
למידת מכונה סרטונים חלק ג
למידת מכונה סרטונים חלק ד
למידת מכונה סרטונים חלק ה
למידת מכונה קורסים מומלצים א
למידת מכונה קורסים מומלצים ב
למידת מכונה ערוצי סלבס
מדעי המחשב בחטיבת הביניים



סרטוני הדרכה

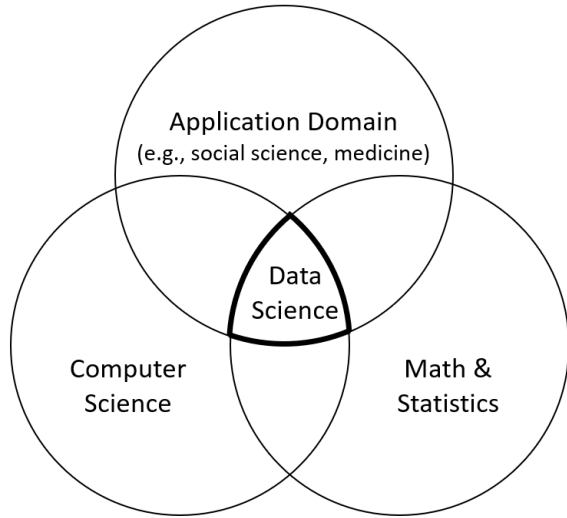
- [מבוא למדעי הנתונים ולמידת מכונה \(30%\) - תכנית לימודים לכיתה י - קובי מייק ואריאל](#)
- [למידה עמוקה - תכנית לימודים לכיתות י"א ו"ב - קובי מייק ואריאל](#)

חומרי לימוד מורים

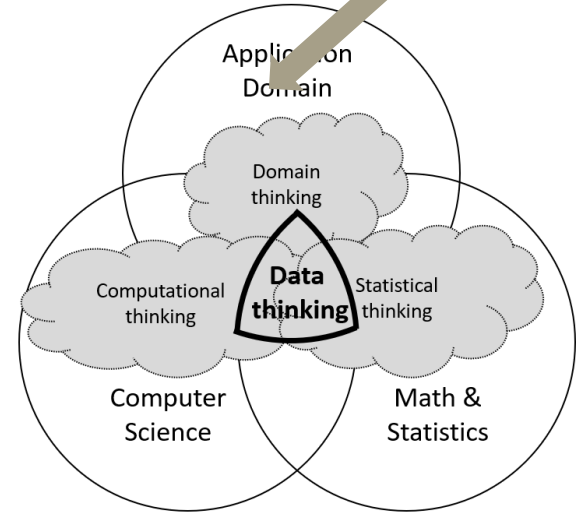
- [אריאל](#)
- [קובי מייק](#)



Data science and data thinking



(a) Data science integrates computer science, mathematics, and statistics, and a real-world domain.



(b) Data thinking integrates computational thinking, statistical thinking, and domain thinking.

Domain thinking

- ▶ The consideration of the context, **the application domain from which data is taken**, is crucial in data science problem solving processes.
- ▶ We saw: The *domain neglect cognitive bias* and.
- ▶ The **interdisciplinarity** of data science is highlighted:
 - the application domain together & the other disciplines (CS, statistics and math)
 - Important in:
 - Evaluation of the performance of an algorithm
 - Ethics
 - More?

אתיקה

מה זו אתיקה? ▶

מדוע צריך אתיקה? ▶

האם קהיליית מדעני.ות הנתונים צריכה אתיקה? ▶

– אם כן: על אלו עקרונות יש לבסס את אתיקה זו?

האם קהיליית מורי.ות מדעי הנתונים צריכה אתיקה? ▶

– אם כן: על אלו עקרונות יש לבסס את אתיקה זו?

אתיקה (על-פי אסא כשר)



מקצועיות נתנת לתיאור כהיררכיה:

- גוף ידע שיטתי
- מיומנות פתרון בעיות הקשורות במקצוע (ארגז כלים)
- חובת ההתעדכנות (היות והידע והמיומנויות מתקדמים)

— הבנה

- היכולת לנמק את הפעילות המקצועית, לענות על שאלות "למה".
- חשיבות ההבנה באה לידי ביטוי בעת פתרון בעיות כאשר לא נמצא כלי מתאים בארגז הכלים.

— אתיקה

- הבנת המקצוע, מהות המקצוע.
- קובעת מהו הראוי – רף עליון להתנהגות (אין באתיקה חוקים במובן של עונשים).

Ethics

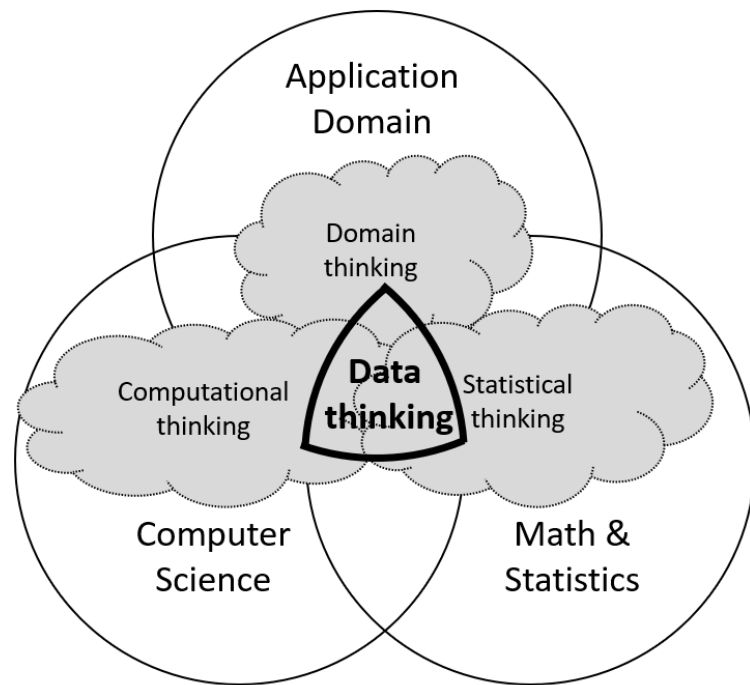
- In addition to each domain ethics (medicine, lawyers, educators,...), from which data is taken, **data science codes of ethics exists.**

הכנת מצגת משותפת: אתיקה של מדעי הנתונים ▶

<https://docs.google.com/presentation/d/1i-r0uLpMu-TavBich1I13O4MYIv8vvBtZnW8lJ5oBo0/edit?usp=sharing>

Summary: Data Thinking

- ▶ The **interdisciplinary** and **multifaceted nature** of data science:
 - Different modes of thinking contributes to data thinking – the mode of thinking required for doing meaningful data science
 - The importance attributed to the application domain in the data science workflow and data science education
 - Algorithm-data duality



Data thinking integrates computational thinking, statistics and math thinking, and domain thinking.