חשיפה למחקר
המחלקה להנדסת תעשייה וניהול

רון הירשפרונג

Ariel Cyber Innovation Center

CHARM Lab
Collaborative Human-Agent Relationship Management

אוניברסיטת אריאל בשומרון

ה"פילוסופיה" שמאחורי המוטיבציה המחקרית

- אדם אינו יכול להזיז אבן כבדה שהונחה במקום מסוים ע"י מכונה.
  להשגת מטרה זו הוא נדרש להשתמש בשירותיה של מכונה אחרת.

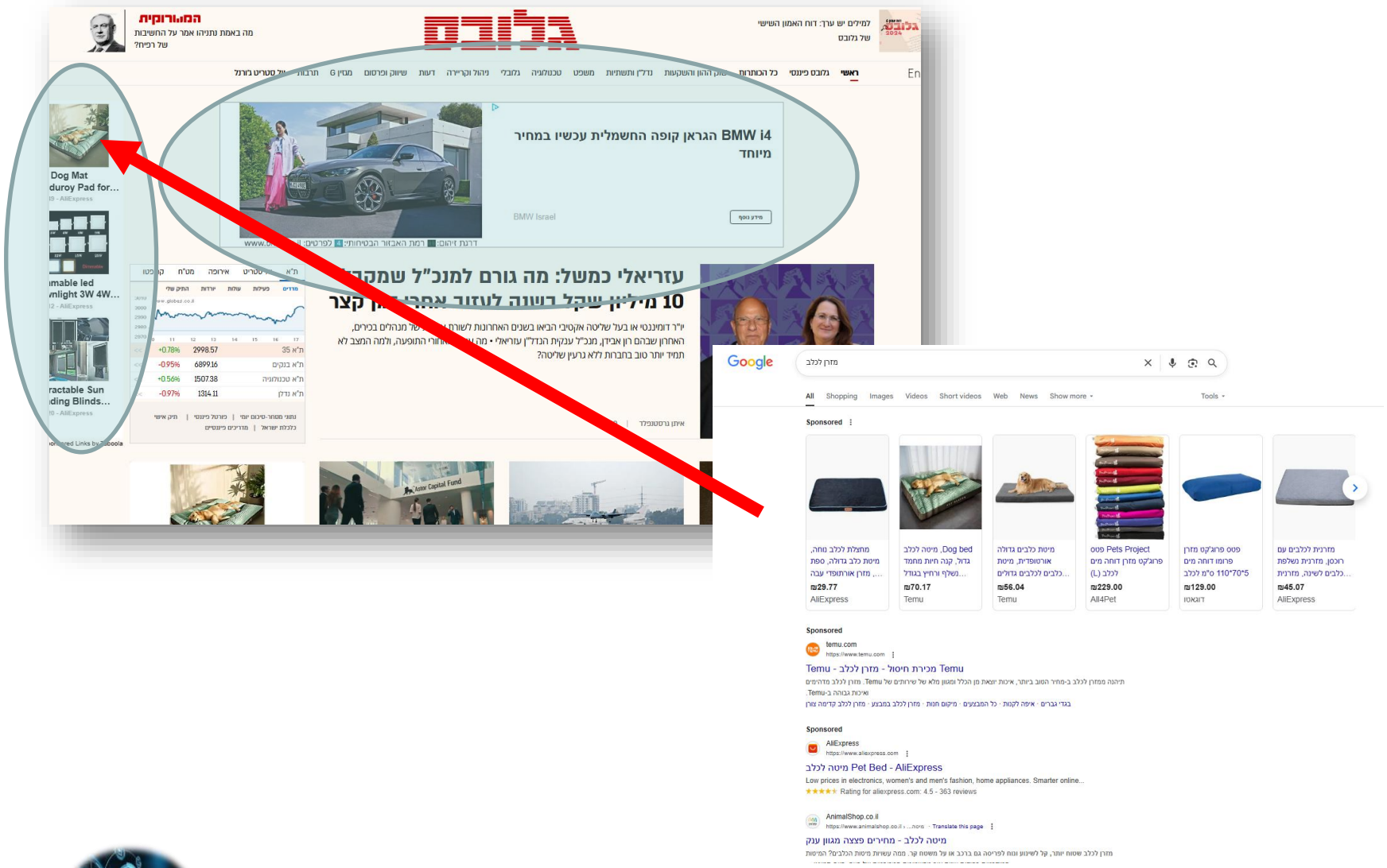
- פרטיות היא סוגייה מרכזית, ונחשבת בעולם המערבי ל-Human Right.

- כמשל: כיוון שהפרטיות מופרת בעיקר ע"י מכונות (AI) ⇦ נדרש שימוש בעוצמה זהה, קרי AI, לשמירת הפרטיות.

- ניתן להגיד שהמחקרים שלי עוסקים ב:

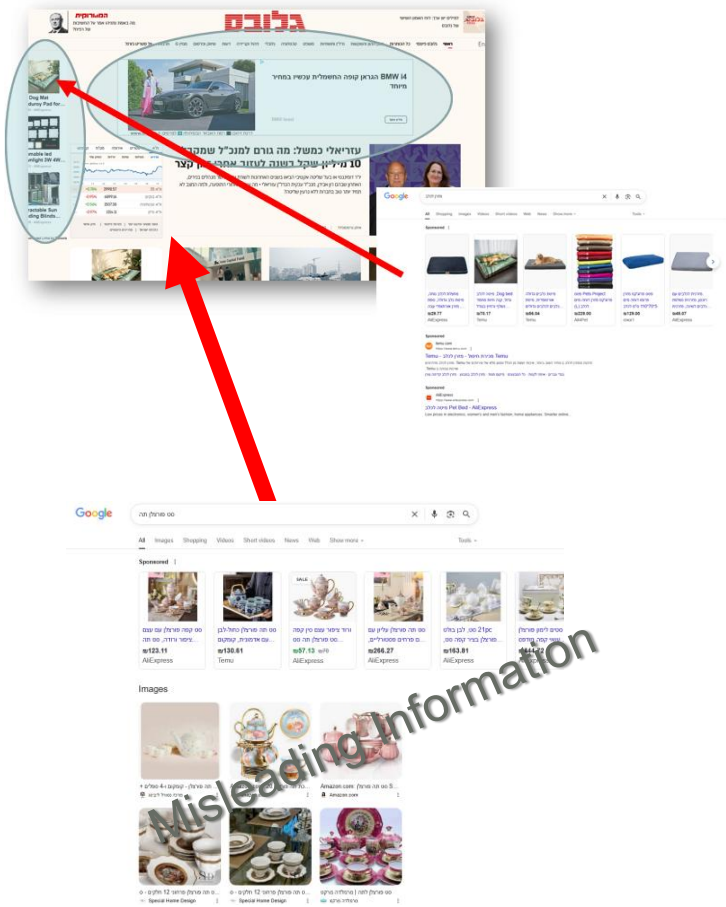
  Harnessing AI to defeat AI

הגנת פרטיות פרואקטיבית



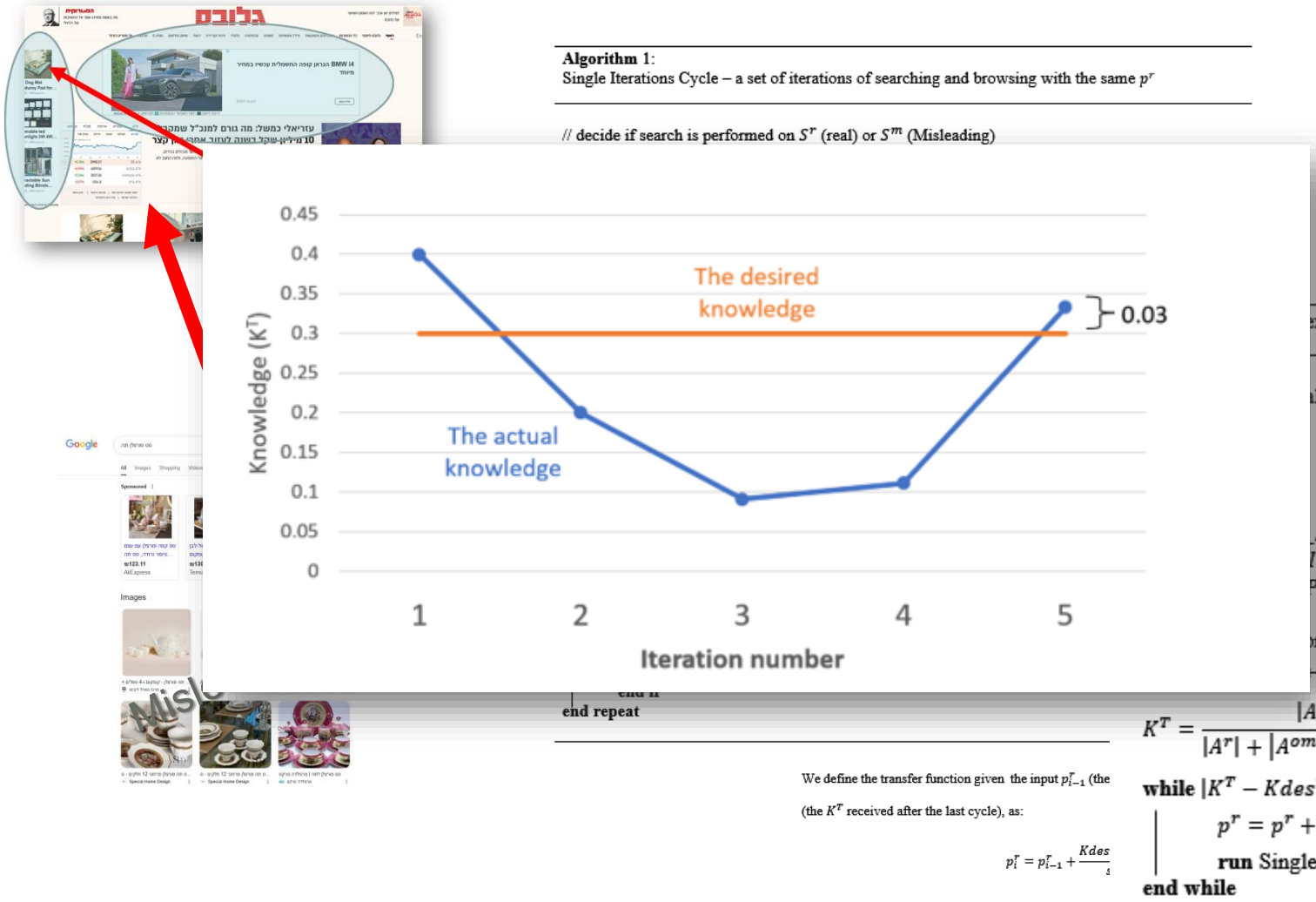
הגנת פרטיות פרואקטיבית

❖ פתרון אפשרי: Anonymization



CONTROL

הגנת פרטיות פרואקטיבית



Algorithm 1:
Single Iterations Cycle – a set of iterations of searching and browsing with the same p^r

// decide if search is performed on S^r (real) or S^m (Misleading)

external process

lization

$\_Pfirst$
$Initial\_DecayFactor$
$Prem$

ns Cycle
– evaluate knowledge()

$$K^T = \frac{|A^r|}{|A^r| + |A^{om^+}| + |A^{om^-}|}$$

while $|K^T - Kdes^T| > TR$

$$p^r = p^r + \frac{Kdes^T - K^T}{step}$$

run Single Iterations Cycle

end while

end repeat

We define the transfer function given the input p^r_{i-1} (the

(the K^T received after the last cycle), as:

$$p^r_i = p^r_{i-1} + \frac{Kdes}{s}$$

מוזיקה כמזהה אישי

"Adon Olam"
https://www.youtube.com/embed/d29qSLUstxw?enablejsapi=1&wmode=opaque&autoplay=0

Makes me happy on many levels - the soulful, minor melody of the verses moves to an upbeat major chorus of hope and appreciation; the words can fit almost any other melody; reminds me of great times with my kids and at shul

danceability	0.438
acousticness	0.763
energy	0.382
instrumentalness	0.00301
liveness	0.112
loudness	-9.3
speechiness	0.0288
tempo	149.967
valence	0.185
key	9
mode	0

anger	0.000956
disgust	0.000689
fear	0.000212
joy	0.98506
neutral	0.005446
sadness	0.00452
surprise	0.003117

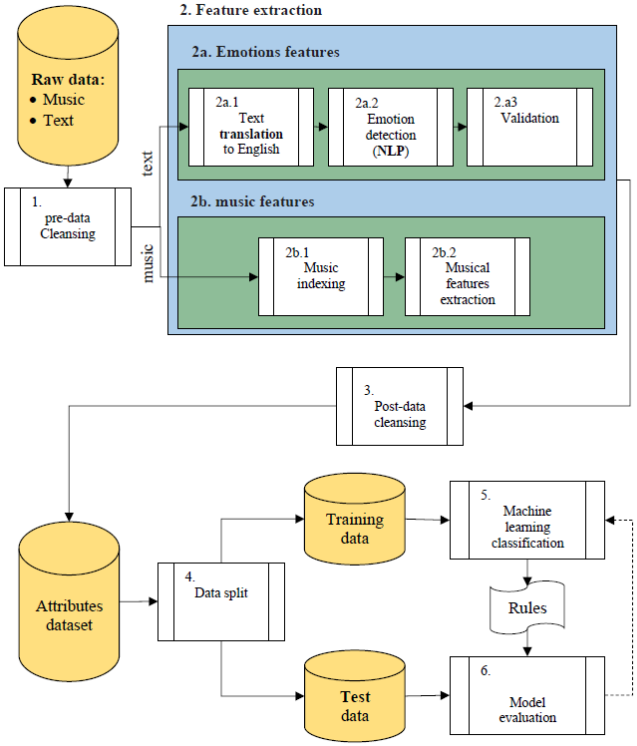
Musical features

- YouTube ID extraction
- SOPTIFY API

Emotions

- text translation
- NLP (DistilRoBERTa)
- Manual validation

מוזיקה כמזהה אישי



- First emotion prediction:

  Accuracy = 0.575 (baseline $= \frac{1}{7} = \sim 14.3\%$)

- One of the two leading emotions prediction:

  Accuracy = 0.751 (baseline $= \frac{12}{42} = \sim 28.6\%$)

ערך זליגת מידע ברשתות חברתיות

Information Leakage

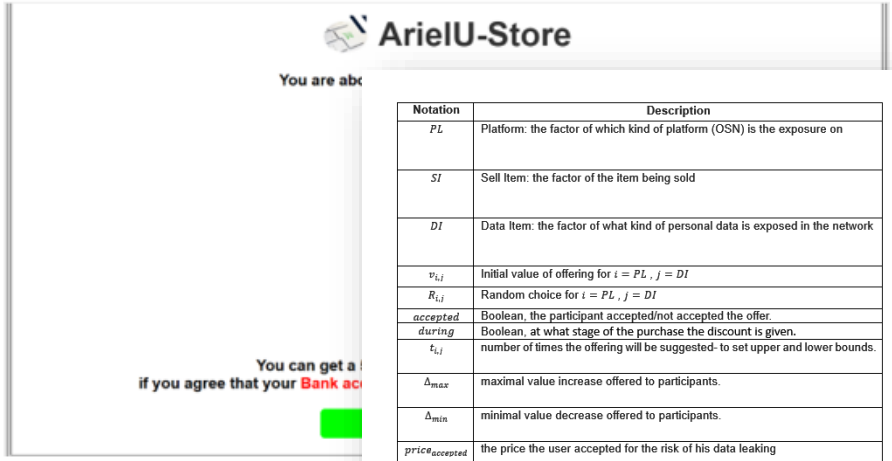
BENEFITS

Trade-off

COSTS

How much? (in tangible values like $)

ערך זליגת מידע ברשתות חברתיות



Notation	Description
PL	Platform: the factor of which kind of platform (OSN) is the exposure on
SI	Sell Item: the factor of the item being sold
DI	Data Item: the factor of what kind of personal data is exposed in the network
$v_{i,j}$	Initial value of offering for $i = PL$, $j = DI$
$R_{i,j}$	Random choice for $i = PL$, $j = DI$
$accepted$	Boolean, the participant accepted/not accepted the offer.
$during$	Boolean, at what stage of the purchase the discount is given.
$t_{i,j}$	number of times the offering will be suggested- to set upper and lower bounds.
Δ_{max}	maximal value increase offered to participants.
Δ_{min}	minimal value decrease offered to participants.
$price_{accepted}$	the price the user accepted for the risk of his data leaking
B_{upper}	upper bound of estimated value of privacy
B_{lower}	lower bound of estimated value of privacy
R_t	random int (0-3) for a minimal 0 times and maximal 3 times of increasing/decreasing offer

The evaluation method

After defining the notation and the matrix of values for the different factors we proceed to define the game's algorithm, that is as follows:

Algorithm 1. $PrivacyValueEstimationGame$
input: $v_{i,j}$, $accepted$, $during$, $t_{i,j}$, Δ_{max}, Δ_{min}, $price_{accepted}$, B_{upper}, B_{lower}, R_t
$B_{lower} \leftarrow 0$
$B_{upper} \leftarrow \infty$
$price_{accepted} \leftarrow 0$
$accepted \leftarrow false$
if $during = false$
  for all the values of SI
  $R_{i,j} \leftarrow$ Random value from the cells $c_{i,j}$ from Matrix (PL X DI)
  Offer a coupon with the value of $v_{i,j}$
    if $accepted = true$
      $price_{accepted} \leftarrow v_{i,j}$
      $B_{upper} \leftarrow price_{accepted}$
    end if
    else
      $B_{lower} \leftarrow v_{i,j}$
      $B_{upper} \leftarrow v_{i,j} + \Delta_{max}$
    end else
  end if
  for all the values of SI
  $R_{i,j} \leftarrow$ Random value from the cells $c_{i,j}$ from Matrix (PL X DI)
  Offer a coupon with the value of $v_{i,j}$
    if $accepted = true$
      $price_{accepted} \leftarrow v_{i,j}$
      $B_{upper} \leftarrow price_{accepted}$
      $t_{i,j} \leftarrow R$
      $index \leftarrow 0$
        while $index < t_{i,j}$
        Offer $v_{i,j} - \frac{index}{t_{i,j}} \Delta_{min}$
        if $accepted = false$:
          $B_{lower} \leftarrow price_{accepted}$
        end if
        else
          $price_{accepted} \leftarrow v_{i,j}$
        end else
        $index \leftarrow index + 1$
      end while
    end if
    else
      $B_{lower} \leftarrow v_{i,j}$
      $B_{upper} \leftarrow v_{i,j} + \Delta_{max}$
      $t_{i,j} \leftarrow R$
      $index \leftarrow 0$
      while $index < t_{i,j}$
        Offer $v_{i,j} + \frac{index}{t_{i,j}} \Delta_{max}$
        if $accepted = true$
          $price_{accepted} \leftarrow v_{i,j} + \frac{index}{t_{i,j}} \Delta_{max}$
          $B_{upper} \leftarrow price_{accepted}$
        end if
        $index \leftarrow index + 1$
      end while
    end else
  end for
return B_{lower} , B_{upper}

ערך זליגת מידע ברשתות חברתיות

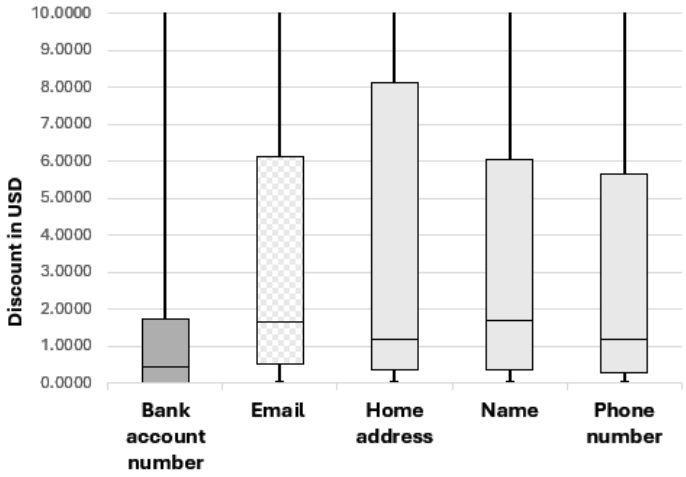


Figure 4. Box Plot results for types of personal data items.

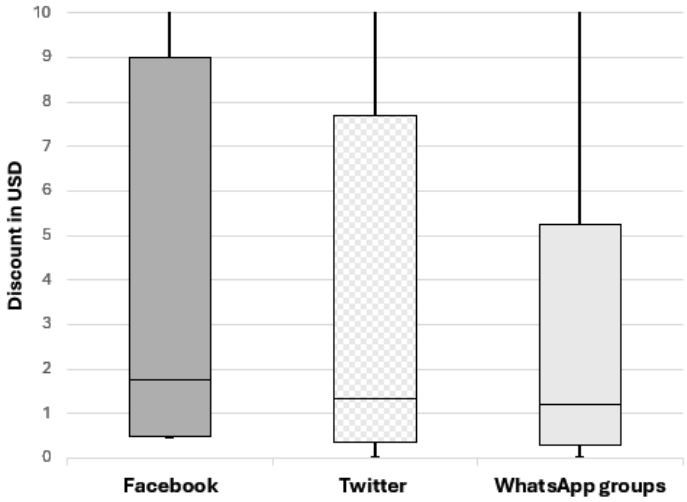


Figure 5. Box Plot results for types of platforms the data is exposed in.

רון הירשפרונג

Ariel Cyber Innovation Center

CHARM Lab
Collaborative Human-Agent Relationship Management

אוניברסיטת אריאל בשומרון