

**როგორ ვიაროთ:  
ხელოვნური ინტელექტისადმი  
ნდობის თუ უნდობლობის გზით?**



**არჩილ ფრანგიშვილი  
თეიმურაზ ნამიჩაიშვილი  
ჟუჟუნა გომიამვილი**

**შემდგენლები:**

**არჩილ ფრანგიშვილი, ოლეგ ნამიჩიშვილი,  
ჟუჟუნა გოგიაშვილი**

**როგორ ვიარათ:  
ხელოვნური ინტელექტისადგი  
დღობის თუ უდღობლობის გზით?**

**ნარკვევი**



**გამომცემლობა „უნივერსალი“  
თბილისი 2023**

## უკ 004.85

მრავალი მაგალითის საფუძველზე საბუთდება, რომ ხელოვნური ინტელექტი არ არსებობს ან უფრო ზუსტად რომ ვთქვათ, ხელოვნური ინტელექტი ჯერ არ არსებობს. ბევრი რამ იქედან, რასაც ხელოვნურ ინტელექტად წარმოგვიჩვენებენ და ნათლავენ, სინამდვილეში მხოლოდ სტატისტიკა და გაუმჭობესებული მანქანური სწავლებაა. ასე რომ დროა ბოლო მოეღოს აკვიატებულ აზრებსა და ფანტაზიებს, რომლებიც დღეს თან ახლავს ხელოვნურ ინტელექტს და მის შესაძლებლობებს.

ნარკვევი განკუთვნილია საქართველოს ტექნიკური უნივერსიტეტის სხვადასხვა პროფილის სპეცილობის ბაკალავრების, მაგისტრანტებისა და დოქტორანტებისთვის, რომლებსაც ღრმა სწავლების გამოყენების სურვილი აქვთ თავიანთი პრობლემების გადასაწყვეტად.

რედაქტორი: პროფესორი **მარიამ ჩხაიძე**

რეცენზენტები: პროფ. **მზია კვიციანი**,

ასოც. პროფ. **ოთარ თავდიშვილი**

© ა. ფრანგიშვილი, ო, ნამიჩიეშვილი, ყ. გოგიაშვილი, 2023

გამომცემლობა „**უნივერსალი**“, 2023

თბილისი, 0186, ა. ჰოლიბაოვსკაიას №4, ☎: 5(99) 17 22 30; 5(99) 33 52 02  
E-mail: universal505@ymail.com; gamomcemlobauniversali@gmail.com

ISBN 978-9941-33-418-4

## სარჩევი

0.	წინასიტყვის ნაცვლად.....	iv
I.	შესავალი.....	1
II.	ხელოვნური ინტელექტი არ არსებობს.....	8
III.	ჩინური ოთახი.....	51
IV.	ადამიანი როგორც ბიოლოგიური რობოტი.	58
V.	ექსპერიმენტის კრიტიკა.....	59
VI.	დაიპყრობს თუ არა სამყაროს მანქანები?...	62
VII.	ა. ნიუელისა და ჰ. საიმონის ჰიპოთეზა.....	62
VIII.	რჩეული მაგალითები.....	64
IX.	ტაუგუს დიპოდი (ორპოლუსა.....	92
X.	დასკვნა.....	95
XI.	ლიტერატურა.....	98

## 0. წინასიტყვის ნაცვლად

ხელოვნური ინტელექტის ინსტრუმენტებისადმი კაცობრიობის საიმედოობისა და ნდობის ხარისხის პრობლემა ამჟამად განსაკუთრებით მწვავედ დგას. ეს დაკავშირებულია იმასთან, რომ ხელოვნურმა ინტელექტმა დაიწყო აქტიურად გავრცელება და შეღწევა ადამიანის საქმიანობის სხვადასხვა სფეროში. ყველა სახელმწიფო მხარს უჭერს და აფინანსებს კვლევებს და ტექნოლოგიურ პროექტებს ამ მიმართულებით. რობოტების შექმნა და ინტელექტუალური სისტემების დანერგვა პროგრამულ პროდუქტებში სამედიცინო მოწყობილობებსა და ტრანსპორტზე სულ უფრო გავრცელებული ხდება. ამასთან დაკავშირებით სავსებით ბუნებრივია შეკითხვები ხელოვნური ინტელექტისადმი ნდობის შესაძლებლობის შესახებ, იმაზეც, თუ რამდენად გამართლებულია ხელოვნური ინტელექტის მიერ მიღებული გადაწყვეტილებები; ვინ არის პასუხისმგებელი ხელოვნური ინტელექტის სისტემის მიერ დაშვებული შეცდომებისთვის, რომლებმაც შეიძლება გამოიწვიოს არა მხოლოდ

ფინანსური, არამედ ადამიანური დანაკარგებიც.

ხელოვნური ინტელექტისადმი ნდობის პრობლემა, მეცნიერთა გარდა, მნიშვნელოვანია ბიზნესისთვისაც, რომელიც უკვე იყენებს ან გეგმავს მსგავსი სისტემების გამოყენებას. ნდობასთან დაკავშირებული სუსტი მხარეების და ყველანაირი პრობლემის აღმოფხვრა ხელს შეუწყობს ბიზნეს-პროცესების გაუმჯობესებას. ნდობის მაღალი დონე — ეს კონკურენტული უპირატესობაა, რადგან, როცა რომელიმე კომპანიას უფრო ენდობიან, ვიდრე კონკურენტებს, ამ კომპანიის ბიზნესი ელვისებური სისწრაფით დაიწყებს ზრდას.

ხელოვნური ინტელექტისადმი ადამიანების ნდობის დონეს მსოფლიოს მრავალ ქვეყანაში ირკვევენ. მაგალითად, «ხელოვნური ინტელექტისადმი ნდობა : ხუთი ქვეყნის მაგალითზე» (Nicole Gillespie, Steve Locke, and Caitlin Curtis, 2021, *Trust in artificial Intelligence: a five country study*, Brisbane, Australia, University of Queensland and KPMG / Klynveld Peat Marwick Goerdeler, 64 pages. <https://doi.org/10.14264/e34bfa3>) — სამეცნიერო

კვლევაა, რომელიც პირველი ღრმა ჩაყვინთვაა ადამიანების მიერ ხელოვნური ინტელექტისადმი ნდობის — და ციფრულ ტექნოლოგიათა გამოყენების მიმართ გაჩენილი მოლოდინების — გაგებაში დედამიწის მრავალადგილზე. ამ გამოკვლევებისთვის გამოკითხული იყო რეპრეზენტატიული ანარჩევი, შედგენილი ამერიკის შეერთებული შტატების, კანადის, გერმანიის, დიდი ბრიტანეთის და ავსტრალიის 1 200 მოქალაქით მაინც (საერთო შერჩევითი ერთობლიობა 6 054 ადამიანს შეადგენდა). ავტორებმა ხუთივე ქვეყანაში გამოავლინეს ოთხი საკვანძო ფაქტორი, რომლებიც განაპირობებს გავლენას მოქალაქეთა ნდობაზე ხელოვნური ინტელექტის სისტემებისადმი :

- რწმენა ხელოვნური ინტელექტის გამოყენების უსაფრთხოების უზრუნველსაყოფად მოქმედი წესების და კანონების ადეკვატურობაში;
- ხელოვნური ინტელექტის გავლენა სამუშაო ადგილებზე;
- ხელოვნური ინტელექტის გაგება;

– ხელოვნური ინტელექტის აღქმული (შეთვისებული) გავლენა საზოგადოებაზე.

ამ ჩამონათვალში — მოქმედი წესებისა და კანონების ადეკვატურობასთან დაკავშირებული — პირველი ფაქტორი ყველაზე ძლიერია.

ზემოხსენებულმა კვლევამ დაგვანახა, რომ ხელოვნური ინტელექტისადმი ნდობა ხუთივე ქვბეყანაში დაბალია და მოქალაქენი, როგორც წესი, სიფრთხილით, არაცალსახად ან უნდობლად ხვდებიან, ეპყრობიან, ექცევიან, ეჭვის თვალით უყურებენ ამ სისტემებს. მოქალაქეთა მხოლოდ მეოთხედი (28%) არის მზად ნდობა გამოუცხადოს მთლიანობაში ხელოვნურ ინტელექტს. ორ მოქალაქეს ხუთიდან არ სურს გაუზიაროს თავისი ინფორმაცია ან მონაცემები ხელოვნური ინტელექტის სისტემას, ხოლო ყოველი მესამე არ ენდობა ხელოვნური ინტელექტის სისტემათა დასკვნებს (ვთქვათ, რეკომენდაციებს ან გადაწყვეტილებებს).

ავტორებს მიაჩნიათ, რომ მთავრობას და ბიზნესს მოეთხოვება შემდეგი ზომები საზოგადოებრივი ნდობის



გასამტკიცებლად ხელოვნური ინტელექტისადმი :

- მოქალაქეთა მოლოდინის გამართლება ხელოვნური ინტელექტის საიმედოობასთან დაკავშირებით;
- ხელოვნური ინტელექტის სამართავად ნორმატიული ბაზის განმტკიცება;
- მოსახლეობის ცნობიერების ამაღლება ხელოვნური ინტელექტის სფეროში.

უმაღლესი დონის ექსპერტთა ევროპულმა ჯგუფმა ხელოვნური ინტელექტის დარგში 2019 წლის აპრილში წარმოადგინა ეთიკის სახელმძღვანელო პრინციპები საიმედო ხელოვნური ინტელექტისთვის. ხსენებული ექსპერტები თვლიან, რომ ხელოვნურ ინტელექტთან საჭიროა მიმდობელ, ნდობით აღსავსე ურთიერთობათა დამყარება და ისეთი ხელოვნური ინტელექტის შექმნა, წახალისება, რომელიც ნდობას იმსახურებს. ამ დოკუმენტის თანახმად, სანდო ხელოვნურ ინტელექტს შემდეგი გარკვეული თვისებები უნდა ახასიათებდეს, სახელდობრ, იგი უნდა იყოს :

- კანონიერი ან უფრო ზუსტად, კანონმორჩილი (ყველა მოქმედი კანონისა და დადგენილების შემსრულებელი, მორჩილი, დამცველი);
  - ეთიკური (ეთიკური პრინციპებისა და ფასეულობათა დაცვით);
  - საიმედო (როგორც ტექნიკური თვალსაზრისით, ასევე სოციალური გარემოს გათვალისწინებით).
- გარდა ამისა, დოკუმენტში დასახელებულია შვიდი საკვანძო მოთხოვნა, რომლებსაც ხელოვნური ინტელექტის სისტემები უნდა შეესაბამებოდეს, რომ შეიძლებოდეს მათი ნდობა :
- ხელოვნური ინტელექტის სისტემები უნდა აფართოებდეს ადამიანების შესაძლებლობებს, აძლევდეს მათ შეგნებულ გადაწყვეტილებათა მიღების საშუალებას და იცავდეს მათ ძირითად უფლებებს;
  - ხელოვნური ინტელექტის სისტემის ტექნიკური საიმედოობა და უსაფრთხოება უნდა იყოს მდგრადი და რეალური;
  - კონფიდენციალობა და მონაცემთა მართვა ;

- გამჭვირვალობა : მონაცემების, სისტემების და ხელოვნური ინტელექტის ბიზნეს-მოდელები უნდა იყოს გამჭვირვალე ;
- მრავალსახეობა, დისკრიმინაციის არარსებობა და სამართლიანობა ;
- საზოგადოებისა და გარემოს კეთილდღეობა ;
- ანგარიშვალდებულება : აუცილებელია ხელოვნური ინტელექტის სისტემებზე და მათ შედეგებზე პასუხისმგებლობისა და ანგარიშვალდებულების უზრუნველყოფი მექანიზმების შექმნა.

მისურის შტატის მეცნიერებისა და ტექნოლოგიების უნივერსიტეტის ბიზნესის და საინფორმაციო ტექნოლოგიების კათედრის გამგე პროფესორი კენ სიაუ მიიჩნევს, რომ ხელოვნური ინტელექტისადმი ორი ტიპის ნდობა არსებობს: თავდაპირველი და მუდმივი (Keng Siau and Weiyu Wang, 2018, *Building Trust in Artificial Intelligence, Machine Learning, and Robotics*, Cutter business technology journal, V. 31, I. 2, Pp. 47-53, Cutter Consortium).

**ცხრილი. ხელოვნური ინტელექტისადმი ნდობის ფორ-  
მირებაზე მოქმედი ფაქტორები**

<b>თავდაპირველი ნდობის ფორმირება</b>	<b>მუდმივი განვითარებადი ნდობა</b>
<p><i>მწარმოებლურობა</i></p> <p>1.წარმომადგენლობა (წარმოდგენა) 2.სახე / აღქმა 3.სხვა მომხმარებელთა გამოცდი- ლები (გამომხმარებლები)</p>	<p><i>მწარმოებლურობა</i></p> <p>1.გამოყენების მოხერხებულობა და საიმედოობა 2.თანაშრომლობა და კომუნიკაცია 3. უსაფრთხოება და კონფიდენცია- ლობის დაცვა 4. ინტერპრეტირებადობა</p>
<p><i>პროცესი</i></p> <p>1. გამჭვირვალობა 2.სასამართლო განხილვის (გარჩე- ვის) შესაძლებლობა</p>	<p><i>მიზანი</i></p> <p>1. ადამიანთა სამუშაო ადგილების ჩანაცვლება 2. მიზნებთან შესაბამისობა</p>
წყარო: [Siau & Wang, 2018, გვერდი 51]	

განვიხილოთ უფრო დაწვრილებით ცხრილში მოცემუ-  
ლი ფაქტორები :

**1. წარმოდგენა.**

წარმოდგენა მნიშვნელოვან როლს თამაშობს ნდობის  
თავდაპირველ აგებაში. თვლიან, რომ, რაც უფრო მეტად

წააგავს რობოტი ადამიანს, მით უფრო ადვილია ადამიანებისთვის ემოციური კავშირის დამყარება მასთან. ძალღი-რობოტი — კიდევ ერთი მაგალითია ხელოვნური ინტელექტის იმ წარმოდგენის, რომლისადმი ადამიანებს უადვილდებათ ნდობის გამოჩენა.

## 2. *სახე / ალექსა.*

სამეცნიერო ფანტასტიკის წიგნებმა და ფილმებმა ხელოვნური ინტელექტის სისტემათა ცუდი სახე შექმნა : ადამიანს შეუძლია იფიქროს, რომ ხელოვნური ინტელექტი უკვე არ ემორჩილება მის კონტროლს და შეუძლია მიაყენოს მას ზიანი. შესაბამისად, მასობრივი ინფორმაციის საშუალებებში, კინოსა და წიგნებში აუცილებელია ხელოვნური ინტელექტის ალტერნატიული გაშუქება. ადამიანებმა უნდა გაიგონ, რომ ხელოვნური ინტელექტი მათი მტერი არ არის, იგი ადამიანების დამხმარე და ხელშემწყობია.

## 3. *სხვა მომხმარებელთა გამოხმაურებები.*

დღეს გამოხმაურებების, გამოძახილებების გავრცელება ძალიან სწრაფად ხდება «Internet» ქსელის წყალობით.

დადებითი გამოხმაურებები ძალიან უწყობს ხელს კოტენციური მომხმარებლების მხრიდან ნდობის ფორმირებას.

#### **4. *გამჭვირვალობა.***

ხელოვნური ინტელექტის პრაქტიკულ გამოყენებებს რომ ენდოს, ადამიანმა საჭიროა გაიგოს, როგორ არის ისინი დაპროგრამებული და რა ფუნქციას შეასრულებს გარკვეულ პირობებში. ეს გამჭვირვალობა მნიშვნელოვანია და ხელოვნურ ინტელექტს უნდა შეეძლოს თავისი ქცევისა და გადაწყვეტილების ახსნაც და გამართლებაც. მანქანური სწავლებისა და ღრმა (სიღრმისეული) სწავლების ერთ-ერთ პრობლემას წარმოადგენს «შავი ყუთი» გადაწყვეტილებათა მიღების პროცესში. თუ ხელოვნური ინტელექტის გამოყენებათა ახსნადობა ცუდია ან საერთოდ არ არსებობს, ეს უარყოფითად მოქმედებს ადამიანთა ნდობაზე ხელოვნური ინტელექტის მიმართ.

#### **5. *გამოცდის შესაძლებლობა.***

ტესტირებადობა ადამიანებისთვის ნიშნავს შესაძლებ-

ლობას მოიპოვონ შეღწევის უფლება ხელოვნური ინტელექტის გამოყენებით პროგრამულ პროდუქტში და მიიღონ მისი მოსინჯვის საშუალება მისი გამოყენების დაწყებამდე. ტესტირება გაგების გაუმჯობესების საშუალებას იძლევა. ამრიგად, პოტენციური მომხმარებლებისთვის ახალი ტექნოლოგიის გამოცდის საშუალების მიცემა ხელს შეუწყობს თავდაპირველი ნდობის უფრო მაღალ დონეს.

***გამოყენების მოხერხებულობა და საიმედოობა.***

მწარმოებლურობა, პირველ რიგში, გულისხმობს ხელოვნური ინტელექტის კომპეტენტურობას ამოცანათა შესრულებაში, ამ ამოცანათა განხორციელებას თანამიმდევრულად და საიმედოდ. ხელოვნური ინტელექტის გამოყენებითი პროდუქტი უნდა იყოს ისეთნაირად შექმნილი, რომ მასთან მუშაობა ადვილად და ინტუიციითაც კი შეიძლებოდეს. არ უნდა წარმოიქმნებოდეს რაიმე მოულოდნელი მოცდენები ან ხანმოკლე მტყუნებები. გამოყენებადობის მოხერხებულობა და საიმედოობა ხელს უწყობს მუდმივ ნდობას.

**7. თანამშრომლობა და კომუნიკაცია.**

თუმცა ხელოვნური ინტელექტის გამოყენებითი პროდუქტების უმრავლესობა იქმნება ამოცანების დამოუკიდებლად შესრულებისთვის, ყველაზე მოსალოდნელი სცენარია ხანმოკლე პერსპექტივაში ის, რომ ადამიანებს მოუწევს მანქანებთან მუშაობა როგორც პარტნიორებთან. იმაზე, თუ რამდენად ადვილად და შეუფერხებლად განხორციელდება თანამშრომლობა და კომუნიკაცია, დამოკიდებულია მუდმივი ნდომა.

**8. თანაზიარობა და კავშირი.**

რაკი ადამიანები — სოციალური არსებებია, მუდმივი ნდობა შეიძლება იყოს გაძლიერებული საზოგადოებრივი საქმიანობით. ძალ-რობოტთან, რომელსაც არ უძნელდება თავისი პატრონის ცნობა და მის მიმართ ერთგულების გამოჩენა, შეიძლება მოჰყრობა როგორც ნამდვილ ძალთან, ასევე ემოციური კავშირისა და ნდობის დამყარება.

**9. უსაფრთხოება და კერძო ცხოვრების დაცვა.**

ეფექტური უშიშროება და მონაცემების დაცულობა —



ორი მნიშვნელოვანი ფაქტორია, რომლებიც გავლენას ახდენს ნდობაზე ტექნოლოგიებისადმი. საეჭვოა, რომ ადამიანებმა გამოიჩინინ ნდობა რაიმე ისეთის მიმართ, რაც ძალიან სარისკოა მუშაობისას. მონაცემთა უსაფრთხოება მნიშვნელოვანია იმიტომ, რომ მანქანური სწავლება ეყრდნობა მონაცემთა დიდ მოცულობებს, რაც პრობლემას უქმნის კონფიდენციალობას.

**10. ინტერპრეტირებადობა.**

მანქანური სწავლების მოდელების უმრავლესობა «შავი ყუთით» მიუწვდომელი და გაუგებარია. ამ პრობლემის გადასაჭრელად აუცილებელია ინტერპრეტირებადი მოდელების აგება და მანქანისთვის მათი დასკვნების ან მოქმედებების ახსნის შესაძლებლობის დატოვება. ამან შეიძლება დახმარება გაუწიოს მომხმარებლებს შედეგების და მათი მიღების პროცესის დასაბუთების გაგებაში.

**11. ადამიანთა სამუშაო ადგილების ჩანაცვლება.**

ხელოვნურმა ინტელექტმა შეიძლება გადააჭარბოს ადამიანურ მწარმოებლურობას მრავალ ასპექტში და ჩანაცვლოს კიდევ ადამიანები მათ სამუშაო ადგილებზე.

ხელოვნური ინტელექტი გააგრძელებს თავისი შესაძლებლობების გაფართოებას და ახალ სფეროებშიც შეიჭრება. შეშფოთება იმის გამო, რომ ხელოვნური ინტელექტი დაიკავებს სამუშაო ადგილებს და ჩანაცვლებს ადამიანებს, ხელს შეუშლის ადამიანთა ნდობის გაღრმავებას ხელოვნური ინტელექტის მიმართ. მაგალითად, იმ ადამიანის თვალში, რომლის სამუშაო ადგილის ჩანაცვლება რეალურად მოხდება, ხელოვნურმა ინტელექტმა სამუდამოდ შეიძლება დაკარგოს ამ პიროვნების ნდობა. დაბალი კვალიფიკაციის, განმეორებად ოპერაციებთან დაკავშირებული და საშიშროების შემცველი სამუშაო ადგილები იმათ რიცხვს მიეკუთვნება, რომლებიც, დიდი ალბათობით, ჩანაცვლებული აღმოჩნდება ხელოვნური ინტელექტის სისტემებით. დაზარალებული მუშაკებისთვის გადამზადებისა და ახალი განათლების ორგანიზებამ შეიძლება შეარბილოს ეს უარყოფითი გავლენა უწყვეტ, მუდმივ ნდობაზე.

**12. მიზნებთან შესაბამისობა.**

ვინაიდან ხელოვნურ ინტელექტს აქვს ის პოტენციალი,

რომ მოახდინოს ადამიანური ინტელექტის დემონ-სტრირება და გადააჭარბოს კიდევ მას, გასაგებია, რომ ადამიანები ხედავენ მასში საშიშროებას, საფრთხეს, მუქარას. ხელოვნური ინტელექტის მიზნები უნდა შეესა-ბამებოდეს მმისთვის დაკისრებულ ამოცანებს და არ გა-მოდიოდეს ამ საზღვრებიდან. მაშინ ნდობა ხელოვნური ინტელექტისადმი იქნება სტაბილური.

*კრიტიკული შენიშვნები ნდობის ცნების გამოყენებაზე*  
არსებობს ალტერნატიული აზრიც იმის შესახებ, რომ ხელოვნური ინტელექტის სისტემებს ადამიანების ნდო-ბა არ ესაჭიროება.

მაგალითად, 2018 წლის 13 დეკემბრის სტატიაში «*No One Should Trust AI*» («არავინ უნდა ენდობოდეს ხე-ლოვნურ ინტელექტს»), რომელიც დიდი ბრიტანეთის ქალაქ ბათის უნივერსიტეტის გამოთვლითი ტექნიკის დეპარტამენტის ეთიკისა და ტექნოლოგიის ასოცირე-ბულმა პროფესორმა დოქტორმა ჯოანა ბრაისონმა (Dr. Joanna Bryson, Associate Professor of Ethics and Technology in the Department of Computing at the University of Bath,

United Kingdom) მოამზადა მასალად გაერთიანებული ერების ორგანიზაციის უნივერსიტეტის (ტოკიო, იაპონია) პოლიტიკურ კვლევათა ცენტრის (Centre for Policy Research, United Nations University, Tokyo, Japan) ინტერნეტ-გამოცემისთვის «ხელოვნური ინტელექტი და გლობალური მართვა-გამგეობა» («AI & Global Governance») <https://cpr.unu.edu/publications/articles/ai-global-governance-no-one-should-trust-ai.html>, იგი წერს, რომ ნდობა — ეს არის ურთიერთობები თანასწორთა შორის, როცა განმდობ მხარეს, რომელმაც არ იცის დანამდვილებით, რას გააკეთებს განდობილი მხარე, მაინც სჯერა მისი ყველა დანაპირების. ხელოვნური ინტელექტი — ეს სისტემათა შექმნის იმ მეთოდების კრებულია, რომლებიც მანქანებს მონაცემთა ერთობლიობის საფუძველზე მოქმედებათა ან ცოდნის დადგენის საშუალებას აძლევს. პროგრამული უზრუნველყოფის შექმნის მხოლოდ სხვა მეთოდი შეიძლება გაიგივდეს ხელოვნურ ინტელექტთან ამ ასპექტში, მაგრამ, ვინაიდან ამ მეთოდ-

დებისთვის უცხოა ცნება «ნდობა» ვინმეს მიმართ, ამიტომ, ბუნებრივია, სრული უაზრობაა მის ანალოგ ხელოვნურ ინტელექტისთვისაც საპასუხოდ ნდობის გამოცხადება.

და რაც უფრო მნიშვნელოვანია, არცერთი ადამიანი არ უნდა ენდობოდეს ხელოვნური ინტელექტის სისტემას, იმიტომ რომ ხელოვნური ინტელექტის საინჟინრო ობიექტა დაპროექტება მკაცრი პასუხისმგებლობის მოთხოვნით შესაძლებელია და სასურველიც. ჩვენ არ გვჭირდება ნდობის გამოცხადება ხელოვნური ინტელექტის სისტემისთვის, შეგვიძლია ვიცოდეთ, რამდენად მოსალოდნელია, რომ იგი შეასრულებს დასმულ ამოცანას და მხოლოდ მას. როცა სისტემას, სადაც ხელოვნური ინტელექტია გამოყენებულია, ზიანი მოაქვს, ჩვენ უნდა ვიცოდეთ, რომ შეგვიძლია იმ ადამიანების პასუხისმგებლობის საკითხი დავაყენოთ რომლებიც ამ სისტემის უკან დგანან.

ბევრი ამტკიცებს, რომ ასეთი პასუხისმგებლობა შეუ-

ძლებელია ხელოვნური ინტელექტის სისტემის შემთხვევაში მისი სირთულისა და იმ ფაქტის გამო, რომ იგი ან აგებულია მანქანური სწავლებაზე, ან ავტონომიურობის გარკვეული ხარისხით ხასიათდება. მაგრამ ვინმეს შეუძლია იმის უარყოფა, რომ ჩვენ საუკუნეებია პასუხს ვთხოვდით და ვსჯიდით ადამიანების მიერ მართვად (მართულ) მრავალ დაწესებულებას, როგორცაა ბანკები ან მთავრობები. ადამიანები ამ დაწესებულებებში ასევე სწავლობენ და ავტონომიით სარგებლობენ, ხოლო მათი ტვინის ფუნქციონირება გაცილებით უფრო მიუწვდომელია და ყოვლად გაგუგებარი, ვიდრე ადამიანის მიერ შეგნებულად შექმნილი და მის მიერ მომსახურებული ნებისმიერი სისტემა. მიუხედავად ამისა, ჩვენ შეგვიძლია დამსახურებისა და ბრალეულობის განსაზღვრა, როცა ეს აუცილებელია, ანგარიშებისა და ჟურნალების შემოღებით, სადაც აღინუსხება ვინ რას აკეთებს და რატომ.

იგივე სამართლიანია ხელოვნური ინტელექტისთვის და არის ამის დამადასტურებელი ფაქტებიც. ვთქვათ,

«უმძლოლო» ავტომობილის მიერ გამოწვეული ყოველი ლეტალური შენთხვევის შესახებ სამყარო ერთი კვირის განმავლობაში იგებს : რას შეიგრძნობდა ავტომობილი, რას თვლიდა იგი სენსორების ჩვენებებად, და რატომ მოქმედებდა სწორედ ისე, როგორც მოქმედებდა და არა სხვანაირად. ავტომობილის მწარმოებელ კომპანიებს კი მიცემენ პასუხისგებაში.

სამწუხაროდ, ხელოვნური ინტელექტის მხოლოდ მცირე ნაწილი შეესაბამება უმძლოლო ავტონობილების სტანდარტებს. ვინაიდან ავტომობილები განსაკუთრებით დამღუპველია და საშიშროების შემცველი, საავტომობილო მრეწველობის რეგულირება კარგად ხდება. მხოლოდ ახლა ვიწყებთ ჩვენ იმის გაგებას, რატომ უნდა ხდებოდეს პროგრამული სისტემების, ხელოვნური ინტელექტით ან მის გარეშე, ასეთივე კარგი რეგულირება. დღეს უკვე ვიცით, რომ უზარმაზარი ზიანი შეიძლება იყოს მიყენებული ქვეყნისთვის, როცა, მაგალითად, უცხო ქვეყნების მთავრობები ერევა

მოახლოებულ არჩევნებში სოციალური ქსელების მეშვეობით. იგივე ძირითადი სტანდარტები, რომლებიც გამოიყენება ვაჭრობისა და წარმოების სხვა დარგებში, შეიძლება იყოს გამოყენებული პროგრამულ სისტემათა სარეგულირებლად. ისევე, როგორც ნებისმიერი სხვა ნაწარმოები პროდუქტის შემთხვევაში, პროდუქტის მწარმოებელი ან მფლობელი / ოპერატორი პასუხს უნდა აგებდეს ნებისმიერი მიყენებული ზიანისთვის. წინააღმდეგ შემთხვევაში ბოროტმოქმედი შეეცდება თავი დააღწიოს პასუხისმგებლობას მის მიერ შექმნილ პროგრამულ სისტემასთან დაკავშირებით და ბრალი დასდოს სისტემის ისეთ მახასიათებლებს, როგორიცაა ავტონომიურობა ან გონივრულობა.

პროგრამული უზრუნველყოფის დამპროექტებელ ორგანიზაციებს უნდა შეეძლოს სათანადო სიფრთხილით მოქმედება ამ პროგრამული უზრუნველყოფის შექმნისას, რეგისტრაციისთვის შესაბამისი სტანდარტების გამოყენების ჩათვლით : რა კოდია დაწერილი (ასევე როდის, რატომ და ვის მიერ), რომელი პროგრამული



უზრუნველყოფა და მონაცემთა ბიბლიოთეკები გამოიყენება, როგორც მოწყობილობა გამოიყენება სისტემის შექმნისას. ორგანიზაციებს ევალება შესაბამისი ტესტირების ჩატარება და დაკუმენტირება პროგრამული უზრუნველყოფის გამოშვების წინ, ასევე მონიტორინგისა და მომსახურების განხორციელება კოდის გამოყენების პროცესში. ისევე, როგორც სხვა დარგში, მან პასუხი უნდა აგოს, თუ ვერ დაამტკიცებს საჭირო წინდახედულებას. ეს პროცედურები უზრუნველყოფს უფრო უსაფრთხო და სტაბილური პროგრამული სისტემების შექმნას, ინტელექტუალურის თუ ჩვეულებრივის.

საბოლოო ჯამში, ასეთი სტანდარტების დაცვის უზრუნველყოფა უნდა ხორციელდებოდეს ადგილობრივი კანონდამცველებით და არა მხოლოდ სპეციალიზებული მარეგულირებელი ორგანოების მიერ ადგილობრივ, ნაციონალურ და ტრანსნაციონალურ დონეზე. ტრანსნაციონალურ კოორდინაციაში რეალური აუცი-

ლებლობა არსებობს, ვინაიდან ხელოვნური ინტელექტის სისტემათა დამპროექტებელი მრავალი კორპორაცია ეროვნული საზღვრების მიღმა მოქმედებს. მაგალითად, ევროკავშირს მნიშვნელოვანი პრეცედენტი აქვს შექმნილი ამ თვალსაზრისით. განსაკუთრებით საინტერესოა სტანდარტებთან დაკავშირებულ იმ შეთანხმებათა ევროპული მოდელი, რომლებიც შეიძლება იყოს ხელმოწერილი ქვეყნებს შორის და შემდეგ განხორციელებული მათ მიერ. შესაძლოა, რომ მომავალში სახელმწიფოები ასეთ ხელშეკრულებებს სულ უფრო ხშირად დადებს ტრანსნაციონალურ კორპორაციებთან. მნიშვნელოვანია, რომ ხელოვნური ინტელექტის სისტემა არასოდეს იყოს წარმოდგენილი პასუხისმგებელ ან დამოუკიდებელ იურიდიულ პირად. სამართლებრივი შეკავება ეფუძნება ადამიანურ მახასიათებლებს — ტანჯვას, რომელსაც ჩვენ განვიცდით, როცა ვკარგავთ სტატუსს, თავისუფლებას, საკუთრებას, ქონებას და ასე შემდეგ. დღეს არ არსებობს ხელოვნური ინტელექტის ისეთი სისტემის შექმნის ხერხი, რომელიც თავად იქნება

შეპყრობილი ასეთი შეკავებით, ყოველ შემთხვევაში ისე არა, როგორც ადამიანი ან თუნდაც სხვა სოციალური ცხოველი. ხელოვნური ინტელექტი იურიდიულ პირად მიჩნევისას გახდება ყველაზე დიდი ცრუმაგიერი კომპანია. ცნობილია, რომ მრავალ იურუსდიქციაში იურიდიული პირის კონცეფცია შეიძლება იყოს ზედმეტად გაფართოებული და დიდი ალბათობით მოსალოდნელია, რომ კორუფციონერები მიზნად დაისახავენ მის გავრცელებას ხელოვნურ ინტელექტზეც, რათა აიცილონ თავიანთი მოვალეობები საზოგადოების წინაშე.

ჩვენ მხოლოდ ახლა — უდიდესი უთანასწორობის, საზოგადოებათა გამთიშველი, მათი მართვის გართულების გამომწვევი პოლიტიკური პოლარიზაციის პერიოდში — ვიწყებთ ხელოვნური ინტელექტის ადამიანთა ცხოვრებაში შეჭრის მასშტაბების გაცნობიერებას. საჭიროა ხელოვნურ ინტელექტზე კონცენტრირება როგორც შემეცნებისა და კომუნიკაციის ინსტრუმენტზე და არ ჩავიციკლოთ ნარცისულ სწრაფვებზე. ხელოვნური

ინტელექტი — ის სუბსტანცია არ არის, რომელსაც შეიძლება ეწოდოს. ეს პროგრამული უზრუნველყოფის დაპროექტების მხოლოდ მეთოდთა სიმრავლეა და ამ მეთოდების საშუალებით უნდა ავამაღლოთ ჩვენი ინსტიტუტებისა და საკუთარი თავის საიმედოობა.

ამავე აზრისაა მარკ რაიანი (Mark Ryan, 2020, *In AI we trust: Ethics, artificial intelligence, and reliability* // Science and Engineering Ethics. V. 26. I. 5. Pp. 2749–2767, Published 10 June 2020, <https://doi.org/10.1007/s11948-020-00228-y>), სტოკჰოლმის სამეფო ტექნოლოგიური ინსტიტუტის ფილოსოფიის განყოფილების პროფესორი (The Division of Philosophy, KTH/Kungliga Tekniska högskolan Royal Institute of Technology, Stockholm, Sweden). ამ ავტორის შეხედულებით, ნდობა, როგორც ცნება, დაუშვებელია იყოს ნახმარი ხელოვნური ინტელექტის მიმართ. ნდობა გულისხმობს, რომ ერთმა შეიძლება მეორეს დაენდოს, ერთმა შეიძლება მეორის იმედი იქონიოს, თუ ეს უკანასკნელი განმდობის მიმართ *მოქმედებს კეთილი ნებით*. სწორედ ამ მთავარი მიზეზის გამო ადამიანის

მიერ შექმნილი ობიექტი, მაგალითად, ხელოვნური ინტელექტის სისტემა, შეიძლება იყოს საიმედო, მაგრამ არ იმსახურებდეს ნდობას. ნდობასთან მიმართებაში ყველამ პასუხისმგებლობის თავისი წილი უნდა ზიდოს, რისი გაკეთებაც ხელოვნურ ინტელექტს არ შეუძლია, რადგან იგი მეთოდებისა და ალგორითმების კრებულია მხოლოდ.

ამიტომ მარკ რაიანი ურჩევს ადამიანებს ყურადღება გაამახვილონ ხელოვნური ინტელექტის საიმედოობაზე, ხანგამძლეობაზე და არა მის მიმართ ნდობის ხარისხზე იმის გათვალისწინებით, რომ საიმედო ხელოვნური ინტელექტი პასუხისმგებლობის ტვირთს აზიდინებს მათ, ვინც დააპროექტა, დანერგა და იყენებს ამ ტექნოლოგიებს, მაგრამ არ გადააქვს ეს პასუხისმგებლობა სისტემაზე.

ხსენებულ ავტორებს შეიძლება დავეთანხმოთ იმაში, რომ მნიშვნელოვანია არასოდეს ხდებოდეს ხელოვნური ინტელექტის სისტემათა წარმოსახვა პასუხისმგებელ

ან იურიდიულ პირებად. მაგრამ უნდობლობა ხელოვნური ინტელექტის სისტემებისადმი აღიმართება დაბრკოლებად მათი განვითარებისა და გამოყენების წინაშე, როგორც ამას შენიშნავენ ანდრეა ფერარო და მისი თანაავტორები (Andrea Ferrario, Michele Loi & Eleonora Viganò, 2020, *In AI we trust incrementally: A multi-layer model of trust to analyze human-artificial intelligence interactions* // *Philosophy & Technology*. V. 33, I. 3, Pp. 523–539, <https://doi.org/10.1007/s13347-019-00378-3> ). ამიტომ, მათი აზრით, ნდობა უნდა იზრდებოდეს მრავალდონიანი მოდელის შესაბამისად ადამიანისა და ხელოვნური ინტელექტის თანამოქმედების გასაანალიზებლად, მიუხედავად იმისა, რომ ხელოვნური ინტელექტი მხოლოდ ტექნოლოგია არის და არა ადამიანი.

### **დასკვნითი შენიშვნები**

არსებული კვლევების ანალიზიდან შეიძლება გამოიყოს ნდობის უზრუნველყოფის ძირითადი ხერხები ხელოვნური ინტელექტის სისტემათა გარკვეული კატეგორიის პოტენციური მომხმარებლების მხრიდან. მაგალითად,

რაკი ნდობამ შეიძლება უზრუნველყოს უფრო მაღალი კონკურენტუნარიანობა ბაზარზე, ამიტომ ბიზნესისთვის მნიშვნელოვანია ყურადღების გამახილვება შემდეგ ფაქტორებზე :

1. ხელშესახებობა.

ჩატარებულია კვლევები ადასტურებს წარმოდგენას იმის შესახებ, რომ ადამიანის მსგავსი (ანთროპოიდის, ჰომინიდის) ან ცხოველის ვიზუალური სახის მიცემა ხელოვნური ინტელექტის სისტემისათვის მკვეთრად ზრდის ამ სისტემისადმი ნდობას მსგავსი ვიზუალიზაციის არარსებობის შემთხვევასთან შედარებით.

2. გამჭვირვალობა.

მომხმარებლებისგან გადაჭარბებულად დიდი და არარეალური მოლოდინების შესამცირებლად ერთ-ერთი ხერხია — ხელოვნური ინტელექტის ფუნქციონალობის შესახებ ახსნის უზრუნველყოფა მისაწვდომი ფორმით. თუ მომხმარებელს ესმის, როგორ მივიდა ხელოვნური ინტელექტის სისტემა გარკვეულ დასკვნამდე და რისი

გაანალიზება მოუხდა მას ამისთვის, მომხმარებელი შეიძლება დარწმუნებული იყოს მასში და, მაშასადამე, — ნდობითაც აღივსოს მის მიმართ.

3. ნდობა ვირტუალური ხელოვნური ინტელექტისადმი.

ვირტუალური აგენტი ხელოვნური ინტელექტის მხარდაჭერით წარმოდგენილია ფიზიკურად (ესე იგი არ არის პირველი პუნქტი «ხელშესახეობა»), მაგრამ ჩატბოტს ან ვირტუალურ თანაშემწეს შეიძლება გააჩნდეს საკუთარი ვირტუალური წარმოდგენა. ხელოვნური ინტელექტი შეიძლება არსებობდეს ნებისმიერ ელექტრონულ მოწყობილობაზე და გააჩნდეს ისეთი ფუნქციები, როგორცაა ავტარი, სახე, სხეული, ხმა, ანიმირებული მოძრაობები ან პროექციებიც კი რეალურ სამყაროში. ყველაფერ ამას შეუძლია მომხმარებლების ნდობის ხარისხის გაზრდა.

4. საიმედოობა.

საიმედოობა მნიშვნელოვან როლს თამაშობს მომხმარებელთა ნდობაში. გამოკვლევებმა, რომლებიც ნდობის იმ



დონეთა შემოწმებას ახდენდა, რომლებიც განპირობებულია ხელოვნური ინტელექტის სიზუსტით და წარუმატებლობებით, გამოავლინა სტატულური პატერნი (ინგლ. pattern — ნიმუში, შაბლონი, თარგი), რომელიც მიუთითებს იმაზე, რომ ჩაშენებული ხელოვნური ინტელექტის შეცდომები აზიანებს ნდობას. შეცდომა გავლენას ახდენს ნდობაზე ისეთნაირად, რომ ამის გამოსწორება ძალიან ძნელი ხდება. ამიტომ უნდა ხდებოდეს ხელოვნური ინტელექტის სისტემის საგულდაგულო ტესტირება, რეპრეზენტატიული მწვრთნელი (დამსწავლელი) ანარჩევის ფორმირება, ასევე ხელოვნური ინტელექტის სიმტკიცის (ურყეობის, მტკიცე რწმენის, გულდაჯერებულობის) სკალის შექმნა გამოტანილი გადაწყვეტილებისთვის, რათა ხელოვნური ინტელექტის მცირე დაჯერებულობისას მიღებული გადაწყვეტილების სისწორეში, საკითხი გადაეცემოდეს ადამიანს.

ემოციური ნდობა ჩაშენებული ხელოვნური ინტელექტის მიმართ შეიძლება იყოს აგებული ტექნოლოგიის

რეპუტაციისა და მასთან დაკავშირებულ ორგანიზაცია-  
თა რეპუტაციების საფუძველზე. თუ მომხმარებელი  
ენდობა ხელოვნური ინტელექტის დამპროექტებელ  
ორგანიზაციას, მაშინ მას აქვს ნდობის დიდი თავდა-  
პირველი მარაგი ტექნოლოგიისადმი.

5. მოცემულ ამოცანებთან შესაბამობა.

ხელოვნური ინტელექტი არ უნდა გამოდიოდეს მისგან  
მოთხოვნილ ამოცანათა ფარგლებიდან. მომხმარებელს  
არ სურს და ამიტომ არც უნდა იღებდეს აუცილებელზე  
მეტ ინფორმაციას, ვინაიდან ამან შეიძლება შეამციროს  
მისი ნდობა სისტემისადმი. შესაბამისად, დაპროექტე-  
ბისას აუცილებელია კორექტული შეზღუდვების დადე-  
ბა ხელოვნური ინტელექტის სისტემის გამოყენებათა  
არეზე.

6. გამოხმინება (აზრის გამოთქმა).

აუცილებელია იმ პირთა მოტივირება, რომლებსაც  
მოსწონთ პროდუქტი, უნდა ხდებოდეს ამ პროდუქტის  
დადებითი მოხსენიება და სწრაფი რეაგირება თითოე-  
ულ უარყოფით გამოხმინებაზე. ამრიგად, დადებით

გამოხმინებათა დიდი რაოდენობა ჩამოაყალიბებს პო-  
ტენციურ მომხმარებელთა ნდობის საწყის დონეს.

7. უსაფრთხოება და კონფიდენციალობა.

მომხმარებელმა უნდა იცოდეს, რომ მისი პირადი მონა-  
ცემები სრულ უსაფრთხოებაშია, გასაიდუმლოებულია,  
კონფიდენციალობა, გაუსაჯაროება და ფარულობა  
დაცულია კანონით. ამიტომ ბიზნესმა უნდა დასახოს  
ეფექტური ზომები პერსონალური მონაცემების დაცვის  
გასაღრმავებლად. მნიშვნელოვანია სისტემის უსაფრთ-  
ხოების უზრუნველყოფა გატეხვისგან. მომხმარებელთა  
რისკის ხარისხის შემცირებით ხელოვნური ინტელექ-  
ტის გამოყენებისას, ბიზნესი ხელს უწყობს მათი ნდო-  
ბის ზრდას.

ხელოვნური ინტელექტის მიმართ ნდობის ასამაღლებ-  
ლად სახელმწიფომ ყურადღება უნდა მიაქციოს შემდეგ  
ფაქტორებს :

1. ნორმატიული ბაზის განმტკიცება.

მოქალაქეს უნდა ესმოდეს, რომ მის უფლებებს ხელოვ-  
ნური ინტელექტის გამოყენებისას სახელმწიფო იცავს.

ამა თუ იმ შემთხვევისას მისთვის აუცილებელია ზუსტად იცოდეს, სად უნდა მიმართოს და ვინ უნდა იყოს პასუხისმგებელი ხელოვნური ინტელექტის სისტემის მიერ მიღებულ გადაწყვეტილებებზე.

2. მოსახლეობის წიგნიერობის ამაღლება. მსოფლიოში ჩატარებულ კვლევათა საფუძველზე, ცნობილია, რომ გამოკითხულ ადამიანთა მხოლოდ ერთ მესამედს ესმის ხელოვნური ინტელექტის არსი და შეუძლია განმარტოს იგი. განუსაზღვრელობა ბადებს უნდობლობას, ამიტომ აუცილებელია მოქალაქეთა ხელოვნური ინტელექტის მიზნებში და გამოყენების სფეროებში ჩახედულობის ამაღლების პროგრამათა შექმნა

3. ხელოვნური ინტელექტის სახე. ხელოვნური ინტელექტის სისტემებისადმი მოქალაქეთა თავდაპირველი ნდობის ასამაღლებლად აუცილებელია თანდათანობით მოხდეს «ბოროტი ხელოვნური ინტელექტის» ძვალ-რბილში გამჯდარი სახის შეცვლა. ამისათვის საჭიროა ხელოვნური ინტელექტის ჩვენება კინოში, ლიტერატურაში და მასობრივი ინფორმაციის

საშუალებებში როგორც ადამიანის კონტროლს დაქვემდებარებული ტექნოლოგია და არა როგორც «რობოტი-დამპყრობი», «რობოტი-მიმტაცებელი».

დღეს ხელოვნური ინტელექტის სწრაფი, ფართომასშტაბიანი ინვესტირება და დანერგვა ხდება. ამის გათვალისწინებით მნიშვნელოვანი იქნება ხელოვნური ინტელექტის სისტემათა მიმართ — მათი ევოლუციის შესაბამისად — საზოგადოებრივი ნდობისა და მოთხოვნილების რეგულარულად შესწავლა და გადახედვა. ეს უნდა ხდებოდეს ხელოვნური ინტელექტის შესაბამობის უზრუნველსაყოფად ადამიანების მოლოდინთან.

### *მიმართვა მკითხველს*

წინამდებარე ნარკვევში მოცემულია ხელოვნური ინტელექტის განვითარების მხოლოდ რამდენიმე საკვანძო მომენტის აღწერა. მათთან გაცნობის შემდეგ ამ წიგნის მკითხველმა თავად უნდა განსაზღვროს : იაროს ხელოვნური ინტელექტისადმი თავისი ნდობის განუხრელი ზრდის გზით, თუ აირჩიოს ალტერნატიული მიდგომა აქ ჩამოყალიბებული ორი კონცეფციიდან.

## I. შესავალი

ხელოვნური ინტელექტის კონცეფცია აშინებს ადამიანს, რადგან იწვევს მეხსიერებაში ძალაუფლების ბერკეტთან დაკავშირებულ რაღაც შეუცნობ შეგრძნებას.

შიში გამოტანილია აუქციონზე და პირველ „დაბალ დონეზე“ ამ თემასთან დაკავშირებით უკვე მრავალი პუბლიკაცია არსებობს. მეორე, გაცილებით უფრო მნიშვნელოვან დონეზე, ყველაფერი, რაც კი ისახება რისკებისგან თავის დაცვის ან, პირიქით, მათი მართვის საშუალებად ამ შეუცნობი ძალის დამორჩილების მიზნით, იწვევს კიდევ უფრო მეტ დაინტერესებას და საკვირველიც არ უნდა იყოს, რომ ეს კონცეფცია დღეს ყვავის.

პირდაპირ დავიწყოთ თეზით: ხელოვნური ინტელექტი არ არსებობს, იმიტომ რომ იგი არც შეიძლება არსებობდეს, ვიდრე ბუნებრივი ინტელექტი უდიდეს საიდუმლოებად რჩება. ტექნიკის ცნობილი ფრანგი ფილოსოფოსი ბერნარ სტიგლერი (Bernard Stiegler, 1952-2020) იმასაც კი ასაბუთებს, რომ *«ხელოვნური ინტელექტი — ეს ხელოვნური სისულელეა»*. ბუნებრივი ინტელექტის

არა ერთი ასპექტი ცნობილია (მაგალითად, გამოთვლი-  
თი სიმძლავრე და მეხსიერება) და ადვილად შეიძლება  
იყოს მოდელირებული, ჩაყენებული ადამიანის სამსა-  
ხურში. მაგრამ, მანქანისგან განსხვავებით, ხელოვნური  
ინტელექტი არ ფუნქციონირებს როგორც სხეულისა და  
გონებისგან დამოუკიდებელი ავტონომიური მექანიზ-  
მი, რაზეც რომის პაპი ფრანცისკე პირველიც კი ამახვი-  
ლებს ყურადღებას 2020 წლის 23 დეკემბრის საშობაო მი-  
მართვაში [1]. არსებობს ღრმა სტრატეგიული შინაარსის  
მქონე ლოგიკური სამაგიდო თამაში «გო», რომელიც  
ძველ ჩინეთში გაჩნდა, სხვადასხვა შეფასებით სადღაც 2  
– 5 ათასი წლის წინ. მანქანისგან განსხვავებით, გოს მო-  
თამაშე ყოველთვის ერთნაირად ვერ გამოხატავს თავის  
რეაქციას თამაშის პროცესში გაჩენილ ამა თუ იმ სიტუა-  
ციაზე, გამომდინარე დადლილობიდან, ოპტიმიზმის  
გრძნობიდან, მასზე მოწინააღმდეგის მიერ მოხდენილ  
შთაბეჭდილებიდან, რამაც შეიძლება გამოიწვიოს მისი  
მხრიდან უფრო მეტად ან უფრო ნაკლებად თავხედური,  
სარისკო პასუხები, ხშირად ცუდი, ზოგჯერ უკეთესი,

მაგრამ განსხვავებული იმ პასუხისგან, რომელსაც აირჩევდა კომპიუტერი, გამომდინარე სხვადასხვა შესაძლო ვარიანტთან დაკავშირებული მათემატიკური ლოდინის მნიშვნელობიდან.

კულტურა ასევე ახდენს გავლენას ადამიანის ინტელექტის რეაქციაზე, რომლის ნაწილი არაცნობიერია და, შესაძლებელია, არის კიდევ უფრო გაუცნობიერი განზომილებები, ვიდრე კულტურა. რამდენადაც ვიცით, მრავალ ახალგაზრდა ცხოველს შეუძლია მტაცებლის ამოცნობა სხვა ცხოველში, რომელიც მას არასდროს უნახავს.

ამიტომ ავტონომიური ხელოვნური ინტელექტის აჩრდილი, რომელიც საბოლოო ჯამში თავის კონტროლს დაუქვემდებარებს ადამიანს, უნდა იყოს გამორიცხული. ასეთი მოქმედება არ შეიძლება იყოს წმინდა ინტელექტის განვითარების მექანიკური შედეგი, იგი გულისხმობს მანქანის ნებისა და სურვილის არსებობას, მაგრამ დღეისთვის ამ მოტივების წყაროები გამოვლენილი არ არის და ამიტომ არც შეიძლება მივაწეროთ ისინი მანქანას. მანქანას შეიძლება გააჩნდეს მიზანი, რომელიც



მას დაუსახა ადამიანმა, შეიძლება ეხმარებოდეს კიდეც ამ ადამიანს ძალაუფლების გაფართოებაში სხვებზე, მაგრამ მას არ შეუძლია აკეთოს ეს *motu proprio*, რაც ლათინურიდან ითარგმნება ფრაზით — «საკუთარი ინიციატივით». იგი რომის პაპთა განსაკუთრებული *რესკრიპტით ცხადდება*. ამ ფორმულას ხმარობენ პაპის ურთიერთობათა პრაქტიკაში უცხო სახელმწიფოებთან, როცა პაპი არ იძლევა პირდაპირ პასუხს მისთვის გაკეთებულ წინადადებაზე, მაგრამ საქმეს ამის შემდეგ წარმოაჩენს როგორც სრულიად ახალს: თითქოს დღემდე მისთვის არავის შეუთავაზებია რაიმე პროექტი ან წინადადება მოცემულ საკითხზე. ისტორიიდან ისიც არის ცნობილი, რომ პირველი *motu proprio* გამოცემულია 1484 წელს რომის პაპის ინოკენტი მერვის (Innocentius VIII, 1432 – 1492) მიერ.

ასე რომ ლაპარაკია ადამიანის დანამატზე, და არა მის ჩანაცვლებაზე, მაგრამ ევოლუციის ბოლოს ამ თვალსაზრისით ადამიანური ელემენტი, რომელიც არსებითაა თავისთავად, რისკის წინაშე დგება: ეს ისეთივე რისკია,

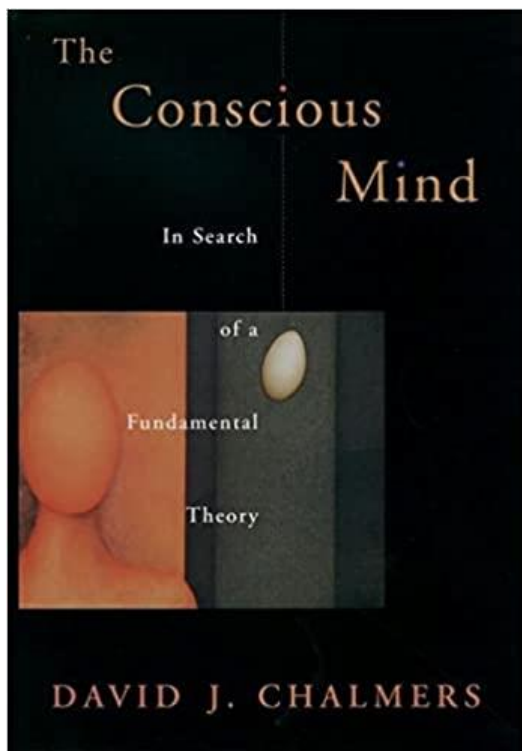
რომ შეწყდება სულის აღქმა ადამიანში რაღაც განცალკევებულობად.

ავსტრალიელი ფილოსოფოსის (ნახ. 1) გენიალური დევიდ ჩალმერსის (David John Chalmers, 1966) მტკიცება იმის შესახებ, რომ მაცივრის თერმოსტატს ცნობიერება აქვს (ან თუნდაც სიმბოლოთა ფიზიკური სისტემის ჰიპოთეზა), შესაძლოა, განსაკუთრებულ ადამიანთა აზროვნების მხოლოდ გაბედულ გაელვებად დარჩეს.



ნახ. 1 დევიდ ჩალმერსი

ეს იდეა დევიდ ჩალმერსმა ჯერ კიდევ 1996 წელს წამოაყენა თავის წიგნში («The Conscious Mind: In Search of a Fundamental Theory»), რომელიც გახდა მეოცე საუკუნის დასასრულის ყველაზე მნიშვნელოვანი ტრაქტატი (ნახ. 2).



ნახ. 2 დევიდ ჩალმერსის წიგნის პირველი გამოცემა (1996)

ჩვენს დროში უკვე არ მოიძებნება სერიოზული ნაშრომი გონებასთან დაკავშირებულ პრობლემაზე, რომელშიც არ მოიპოვებოდეს მისი წიგნის — «შემცნობი ჭკუა» — დამოწმება. ხსენებული წიგნის რუსული თარგმანი (ნახ. 3) მხოლოდ 2013 წელს გამოვიდა და დღემდე რჩება რუსულ ენაზე ნათარგმნ ჩალმერსის ერთადერთ გამოკვლევად.



ნახ. 3 დევიდ ჩალმერსის წიგნის რუსული თარგმანი (2013)

ეს არის წარმტაცი ფილოსოფიური მოთხრობა გონების ღრმა პარადოქსებსა და საიდუმლოებებზე, რომელშიც ვაწყდებით «ნატურალისტური დუალიზმის» დასაბუთების სერიოზულ მცდელობას, გამომდინარე ავტორის შეხედულებიდან გონების არაფიზიკურ ბუნებაზე და მის დამოკიდებულებაზე ტვინში არსებული ფუნქციური სქემების მიმართ. ჩალმერსი ასევე ამტკიცებს, რომ მისი თეორია ახალ პერსპექტივებს გვისახავს კვანტური მექანიკის ინტერპრეტაციისთვის და საშუალებას იძლევა ვილაპარაკოთ შეგნებული რობოტების შესაძლებლობაზე ახლო მომავალში.

## **II. ხელოვნური ინტელექტი არ არსებობს**

უხსოვარი დროის წინად ზევსმა დაავალა ეპიმეთეუსს, იაპეტოსისა და კლიმენას ვაჟიშვილს, გაანაწილოს თვისებები და ღირსებები ყველა ცოცხალ არსებას შორის გაწონასწორებული პროპორციით (ზოგს — კლანჭები და ეშვები სანადიროდ, სხვებს — ჩლიქები, ქერცლი ან ფრთები თავდაცვისთვის). ეპიმეთეუსმა შეასრულა ეს ამოცანა. ყველა ცხოველს მიეცა ძალა თავისი როლის

შესასრულებლად ცხოვრების ამ დიდ სცენაზე. ბუნებას შეეძლო აყვავება, გაფურჩქვნა და უკვდავების შექმნა. მაგრამ ეპიმეთეუსს დაავიწყდა ადამიანი (ძველ ბერძნულ ენაში «Ἐπιμηθεύς» — ლათინურად «Epimetheus» — ნიშნავს იმას, «*ვინც ფიქრს იწყებს მაშინ, როცა ფაქტი უკვე მომხდარია*»). ყველა შესაძლებლობა და ატრიბუტი ამოწურული აღმოჩნდა, ადამიანისთვის არაფერი რჩებოდა. პრომეთემ, ეპიმეთეუსის ძმამ, ესოდენ გაშიშვლებული და სრულიად განიარაღებული ადამიანის ხილვით მიღებული შთაბეჭდილების გავლენით, მოჰპარა ჰეფესტოს და ათენას ყველა ხელოვნების შემქმნელი ცეცხლი და გადააბარა იგი ადამიანებს — ცხოველთა შორის ყველაზე საცოდავ არსებებს. ამ წვრილმან ქურდობას მოჰყვა ღმერთების მკაცრი სასჯელი, ვინაიდან ეპიმეთეუსის მიერ ჩაფიქრებული გაწონასწორებული სამყარო მაშინვე იქცა ერთი სახეობის ტირანიად, ამიერიდან მხოლოდ იგი დაეუფლება ტექნიკას, დაიწყებს სამყაროს მართვას და გაიმარჯვებს ყველა დანარჩენ ცხოველზე.

პლატონის მიერ მის ერთერთ დიალოგში — «კრო-  
ტაგორასი» — მოთხრობილი ეს მითოსი ეპიმეთეუსის  
შესახებ მდიდარია გაკვეთილებით. აწ გარდაცვლილი  
ფრანგი ფილოსოფოსი, რომელსაც ევროპაში XXI საუკუ-  
ნის ერთ-ერთ ყველაზე გავლენიან ფილოსოფოსად მი-  
იჩნევენ, ბერნარ სტიგლერი (Bernard Stiegler, 1952 - 2020)  
ხედავს ამაში შეხსენებას უბედურების, გასაჭირის თაო-  
ბაზე, ანუ, როგორც ბიოლოგები ამბობენ, იმ დღენაკლუ-  
ლობის, უდღეურობის შესახებ, რომელშიც დასაბამიე-  
რად იმყოფება ადამიანი, მაგრამ ტექნოლოგიის ღვთაებ-  
რივი შემოწირულობა (ნაბოძარი) თავდაპირველი ხარ-  
ვეზების და სინაზის გამოსწორების საშუალებას აძლევს  
მას [2]. განგების წყალობით მომხდარი პრომეთეს ჩარე-  
ვით ადამის მოდგმის თანდაყოლილი სისუსტე იქცა  
ჭარბ უპირატესობად, გადამეტებულ ძალად, დამანგრე-  
ველ ძლიერებად სხვა ცოცხალი არსებისთვის და, შესა-  
ძლებელია, თავად მისთვისაც.

ადამიანთა ინსტრუმენტები აფართოებს მათი სხეულე-  
ბის შესაძლებლობებს და მათ აზრებსაც კი ასახავს.

როგორც ანთროპოლოგი ანდრე ლერუა-გურჰანი [3] ამტკიცებს, ყოველი ინსტრუმენტი ახდენს გარკვეული ფუნქციის *ექსტერნალიზაციას*, გარეთ გატანას. თავიდან საქმე ეხებოდა მასალების (ვთქვათ, ჯოხის ან ჩაქუჩის დახმარებით) გამოყენების ყველაზე მარტივ ან ძნელ ხერხებს. შემდეგ ტექნოლოგიათა წვლილი სულ უფრო გაწაფული, დახვეწილი და რევოლუციური ხდებოდა (ენერგეტიკული მანქანების გამოყენებით, რომლებმაც გაზარდა ან ჩაანაცვლა კუნთების უხეში ძალა გაცილებით უფრო ეფექტური ხერხით: ურიკით, ბორბლით, ორთქლის მანქანით, რკინიგზით, ავტომობილით და ა. შ.). მოგვიანებით გაჩნდა სენსორული ხელსაწყოები ანუ ინსტრუმენტები გრძნობისა და აღქმის ორგანოთა შესაძლებლობების დასახვეწად და გასაფართოებლად (ტელესკოპი, მიკროსკოპი, ოპტიკური ტელეგრაფი, ფოტოგრაფია, კინო და ა. შ.). დაბოლოს, გაჩნდა ინფორმაციული სისტემები, ტექნოლოგიათა უკანასკნელი და ყველაზე თანამედროვე თაობა. ამ შემთხვევაში მიზანს უკვე წარმოადგენს კაცობრიობის ინტელექტუალური ფუნქციებისა და შესაძლებლობების გაზრდა და



შეცვლაც კი წიგნების, კალკულატორების, კომპიუტერების, ხისტი დისკოების, ელექტრონული და კომპიუტერული ქსელების და სხვა ფაქტორის დახმარებით.

თუმცა ნუ დავივიწყებთ, რომ უკვე ანბანური დეკომპოზიცია, ლოგიკა, რიტორიკა ან დამახსოვრების (ინფორმაციის შენახვის) ხელოვნება მნიშვნელოვანი არამატერიალური ტექნიკური მიღწევა არის, რომელიც დღესაც გამოიყენება, აადვილებს მრავალათასწლოვანი ძალისხმევის მხარდაჭერასა და სრულყოფას მონაცემების ან ცოდნის ანალიზისა და სინთეზის სფეროში. ეს კი აყალიბებს ადამიანებში *«აზროვნებისა და კლასიფიკაციის»* სიღრმისეულ ინტელექტუალურ ჩვევებს, როგორც ამას ჟორჟ პერეკი (Georges Perec, 1936 - 1982) ამბობს — სახელგანთქმული ფრანგი მწერალი და ბრწყინვალე *ვერბიკრუცისტი* (კროსვორდების კომპოზიტორი, შემდგენელი, ავტორი) [4].

ტექნოლოგიური ორგანოები და მექანიკური ან ელექტრონული *პროთეზები* განუყრელად ახლდა თან ჰომინიზაციას (პროცესს, რომლის მსვლელობისას პრიმატები იქცა ადამიანებად) და უპირველეს ყოვლისა,

ჰუმანიზაციას (პროცესს, რომლის შედეგად ისტორიამდელი (წინარეისტორიული) ადამიანები იქცა ისტორიულ ადამიანებად, პრემოდერნი აზროვნებაში — მოდერნად აზროვნებაში, ველურები — ცივილიზებულებად, ბნელნი — განათლებულებად, ადამიანები — მოქალაქეებად, ბარბაროსული ურდოები — მშვიდობიან საზოგადოებებად, ანარქისტული დაჯგუფებები — სახელმწიფო ორგანიზაციებად).

მაგრამ შეიძლება მოხდეს ისე, რომ ადამიანთა განვითარებადი გაერთიანება და მათი მანქანები მიაღწევს თავის კულმინაციურ წერტილს, შეწყვეტს ცივილიზაციურ პროცესს, დაიწყებს რეგრესირებას, დეჰომინიზაციისა და დეჰუმანიზაციის პროცესების სტიმულირებას, რადგან ტექნოლოგიური აღჭურვილობა გახდება ძალიან ინვაზიური (მოწყობილობიდან მოწყობილობაზე «გადამდები» დაზიანებებით), დამფრთხობი, დამორგუნველი და «დაბერების», «დაძველების» ნიშნების შემცველი.

*შეიძლება, რომ ეს დღე უკვე ნამყოს, წარსულის კუთვნილება იყოს?*

ნებისმიერ შემთხვევაში ბიოლოგიური ევოლუცია მჭიდროდ არის დაკავშირებული ტექნიკურ ევოლუციასთან. ბიოსფერო ძლიერ და სულ უფრო ფართოდ დამოკიდებულია ტექნოსფეროზე. ადამიანის გავლენა ტექნოლოგიაზე პრაქტიკულად სრულია; მაგრამ ტექნოლოგიის გავლენა ადამიანზე ასევე ძალიან მნიშვნელოვანია. ურთიერთობები ადამიანსა და ინსტრუმენტს, არსებასა და მანქანას, ცოცხალსა და რობოტს, ბუნებრივსა და ხელოვნურს შორის თამაშობს მთავარ როლს კაცობრიობის უდიდეს ფილმში, ისტორიაში, საზოგადოებებსა და კულტურებში.

ზოგი ქმნის და იყენებს სხვებს, რომლებიც, თავის მხრივ, ახდენს მათზე უკუქმედებით გავლენას, აყალიბებს მათ აზრებსა და ქცევას. სულ უფრო ძნელი ხდება ადამიანისა და საზოგადოების გამოცალკეება მათი ტექნიკური და ტექნოლოგიური გარემოცვიდან.

წინარე ისტორიული ადამიანის მიერ კაჟის იარაღით ამოკვეთილი კლდის (გამოქვაბულის) პეტროგლიფებიდან დამწერლობის გამოგონებამდე, იოჰანეს გუტენბერგიდან (Johannes Gutenberg, 1397/1400 - 1468) მონაცემთა

მასობრივ დიგიტალიზაციამდე (გაციფვრამდე) — ტექნიკისა და ტექნოლოგიების გრძელი ისტორიაა თავად ისტორიის მსგავსად, რომელიც მის სათავეებთან დგას და კაცობრიობის ევოლუციას უდევს საფუძვლად.

ადამიანის მიერ გამოგონებული ინსტრუმენტები გარდასახავს მას, პირად და პროფესიულ საქმიანობას, კომუნიკაციის საშუალებებს, მოხმარების ხერხებს, ტრანსპორტის სახეობებს ან სამყაროზე დაკვირვებისა და მის შესახებ ფიქრის მეთოდებსაც კი.

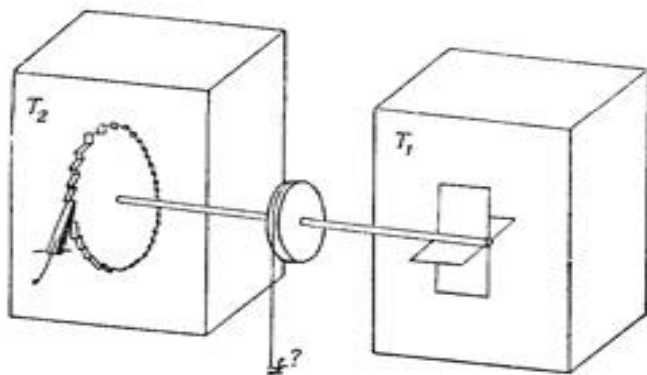
და რაც უფრო მეტი დრო გადის, მით უფრო პრობლემატური ხდება სიახლოვე ადამიანსა და ტექნოლოგიას შორის იქამდის, რომ ჩვენ უკვე არც კი ვიცით, ვისთან გვაქვს საქმე — ერთთან თუ მეორესთან და მაშინ «ტიურინგის ტესტი» გავლილია წარმატებით.

ეს განსაკუთრებით აქტუალურია ურთიერთობისას ხელოვნურ ინტელექტთან, ალგორითმებთან, რობოტებთან, როცა მათი პასუხების რელევანტობის მიღწეული დონე ბუნებრივ ტვინს გვაგონებს.

ტექნოლოგიათა განვითარების ისტორია ხასიათდება თანამიმდევრული გადასვლით სხვადასხვა ზღურბლზე — ე. წ. შეუქცევადობის ხრუტუნაზე (ინგლ. *ratchet wheel*).

ამ უკანასკნელი ტერმინის მეტაფორული გამოყენების მიზეზი ნათელი რომ გახდეს, საკმარისია მივმართოთ ფეინმანის (Richard Philips Feynman, 1918 - 1988) ლექციებს ფიზიკაში [5].

მას აღწერილი აქვს მარტივი მოწყობილობა, რომელიც ლოგიკურად, თითქოს, მხოლოდ ერთი მიმართულებით ბრუნვის საშუალებას აძლევს ღერძს.



ნახ. 4 შეუქცევადობის ხრუტუნა (ratchet wheel) და სასახლეტის (ჭრიალას) შემცველი მანქანა

დავუშვათ, რომ ჭურჭელში არის აირი გარკვეულ  $T$  ტემპერატურაზე, ხოლო შიგნით განთავსებულია ტრიალა (ნახ. 4). ამასთან ერთად დავუშვათ, რომ

$$T_1 = T_2 \equiv T.$$

აირის — სითბურ მოძრაობაში მყოფ — მოლეკულებთან შეჯახების გამო ტრიალა დაიწყებს რყევას. ახლა საკმარისია ღერძის მეორე ბოლოს მივამაგროთ ბორბალი, რომელიც შეძლებს ბრუნვას მხოლოდ ერთი მიმართულებით — *ხრუტუნა ბორბალი საკეტელათი (სასხლეტით, ჭრიალათი)*. სასხლეტი აღკვეთს ტრიალას მცდელობას იბრუნოს ერთ-ერთი მიმართულებით, მაგრამ მის ბრუნვას მეორე მიმართულებით ხელს არ შეუშლის და თავადაც იბრუნებს ამ მიმართულებით. შკივად წოდებული ღარიანი მეორე ბორბალი, რომელიც ხრუტუნასა და ტრიალას შორისაა ღერძზე დამაგრებული და ამ ღარში ძაფია დახვეული, ასევე ნელ-ნელა დაიწყებს ბრუნვას სათანადო მიმართულებით. თუ ძაფზე მოვახერხებთ რწყილის მიბმას, შეიძლება მისი ზევით აწევაც მოხერხდეს და მუშაობის შესრულება!

ცხადია, რომ ჩვენი იდეალიზებული ხრუტუნა ძალზე მარტივია, მაგრამ რაც შეეხება საკეტელას (სასხლეტს, ჭრიალას), მას ესაჭიროება დამატებით ზამზარა. საქმე ისაა, რომ ერთი კბილანას გავლისას სასხლეტი უნდა დაუბრუნდეს წინა მდგომარეობას მომდევნო კბილანაზე, რაც ზამზარის გარეშე ვერ მოხერხდება.

შესაძლებელია ეს? ცხადია, ჩვენ არ ვაპირებთ ამ საკითხის განხილვას, რადგან *«შეუქცევადობის ხრუტუნა»* — აქ მხოლოდ მეტაფორაა ხელოვნური ინტელექტის მიერ მიღწეული პოზიციის შეუქცევადობასთან პარალელის გასავლებლად.

გუტენბერგის მიერ საბეჭდი დაზვის გამოგონება ნიშნავდა შუა საუკუნეთა გადამწერ ბერმონაზონთა გაქრობას, გადაშენებას, ოპტიკური ტელეგრაფისა და რკინიგზათა შექმნამ წარსულს ჩააბარა ცხენები და ეტლები, კომპიუტერებისა და ელექტრონული კალკულატორების გაჩენამ სამგლოვიარო რეკვიემად აქვითინდა საანგარიშო აბაკუსებისთვის და ქალაღზე დაბლაჯნილი თუ დაფხაჭნილი ოპერაციებისთვის; ციფრულმა ფაილებმა და სტრიმინგულმა (ნაკადურმა) მაუწყებლობამ

(ინგლ. *live streaming*), რომელიც მულტიმედიური (ვიდეო და აუდიო) ინფორმაციის ნაკადებს აწოდებს მომხმარებელს და დროის რეალურ რეჟიმში წარმოადგენს მათ, გამოიწვია CD-საკრავებისა და მაღალი სიზუსტით (სისწორით) ბგერის აღმწარმოებელ (Hi-Fi — High fidelity) სისტემათა გამოყენების მკვეთრი შემცირება, არადა თავის დროზე სწორედ მათი გაჩენის გამო გაქრა გრამოფონები (პატეფონები) და მოხდა ვინილის ფირფიტათა ერთგვარი «მარგინალიზაცია» — გამოძევება, თითქმის სრული ამოღება ფართო ხმარებიდან. უეჭველია, რომ ხელოვნური ინტელექტი არის ახალი — და ძალიან მძლავრი — *შეუქცევადობის სასახლეტი, საკეტელა*.

უფრო მეტიც, რევოლუცია ხელოვნური ინტელექტის სფეროში შეიძლება მხოლოდ ახლა იწყებოდეს. ვერავინ იტყვის, როგორი შედეგები დადგება 20, 50 ან 100 წლის შემდეგ. სწორედ ამის გამოა ხელოვნური ინტელექტის თემა ასეთი გატაცების გამომწვევი:

იგი რჩება დიდწილადაც გამოუკვლეველი, შეუსწავლელი და გაურკვეველი.



ამავდროულად სწორედ ეს გარემოება ბადებს ფანტაზიებსა და შიშს. ფუტუროლოგები და ასევე სპეციალისტები ახალი ტექნოლოგიების სფეროში ამტკიცებენ, რომ ხელოვნური ინტელექტი შეცვლის, და ხშირად რადიკალურადაც, მრავალ პრაქტიკულ მიდგომას. მაგრამ მათ მნიშვნელოვნად განსხვავებული აზრი აქვთ იმის თაობაზე, თუ რა მიმართულებით განვითარდება ეს სახენავალი ფორმის პრაქტიკა. პირველ რიგში, ალბათ, ეს შეეხება მკვლევართ: მათი კვლევის ობიექტები არსებითი ხასიათის ევოლუციას განიცდის, ხოლო ამ ობიექტთა შესწავლის საშუალებები მნიშვნელოვნად განახლდება. ხელოვნური ინტელექტი აქტუალური და პერსპექტიული თემაა — ყველა თავის ასპექტში — მრავალი მიმართულების მეცნიერთათვის.

საინტერესოა, რომ *საგანძურად* მიჩნეული წიგნი, რომელიც წარმოადგენს 1971-1994 წლებში თექვსმეტ ტომად გამოცემულ მეცხრამეტე და მეოცე საუკუნეთა ფრანგული ენის ლექსიკონს და 2002 წლიდან თავისუფალ წვდომაშია ინტერნეტში [6], ინტელექტს განსაზღვრავს ასეთნაირად: *«რეალობის ორგანიზაციის მენტალური*

ფუნქცია აზრებში ადამიანთან, და მოქმედებებში ადამიანთან და ცხოველთან». ამავე ლექსიკონის თანახმად, «ინტელექტი არის ფსიქიკური და ფსიქონევროლოგიური ფუნქციების ერთობლიობა, რომლებიც ხელს უწყობს შემეცნებას, საგანთა ბუნების და ფაქტების არსის გაგებას». ამრიგად, საკუთრივ ხელოვნურ ინტელექტს ან საერთოდ არ აქვს ინტელექტი ან მისი ინტელექტი, გონება ძალზე სუსტია.

მაგრამ ინტელექტი, კვლავ ხსენებული წყაროს მიხედვით, ასევე არის «გონებით ჩაწვდომა რაიმეში, მისი გაგება სიღრმისეული ცოდნის სიადვილით». ამ თვალსაზრისით ხელოვნური ინტელექტის გონება (ინტელექტი) ჯერ კიდევ იმყოფება ჩანასახის მდგომარეობაში და სრულფასოვანიც არ არის, რაც უბიძგებს მეცნიერთ წვლილი შეიტანონ მის გაღრმავებასა და დაზუსტებაში. ჩვენი ნარკვევი მიზნად ისახავს მხოლოდ უმნიშვნელო წვლილის შეტანას მის თუნდაც ზედაპირულ გააზრებაში. ყველა ექსპერტი დღეს თანხმდება, რომ ხელოვნური ინტელექტი ვითარდებოდა პრაგმატულად, ვინაიდან ნაპოვნი აღმოჩნდა რეცეპტები, რომლებიც მუშაობს,

თუმცა აქამდის არ არის საკმარისი რაოდენობით კონცეფციები და თეორიები მისი გაგების განსამტკიცებლად და ასახსნელად. ამ თვალსაზრისით ვითარება ინფორმატიკის მდგომარეობის საწინააღმდეგოა, რომელსაც არა ერთი საკმაოდ ძველი კონცეფციის სრულიად ახლებურად რეალიზების საშუალება აქვს ამჟამადაც. ამის საპირისპიროდ, ხელოვნური ინტელექტის ისეთი განმარტებაც კი არ ჩამოყალიბებულა დღემდე, რომელიც ეყრდნობა ამ ფენომენის ამოცნობისა და მოქმედების სფეროს კრიტერიუმებს.

ხელოვნური ინტელექტი *ფანტომად* იმიტომ ეჩვენება ადამიანს, რომ ძალიან ძნელია გაიგო, რას ნიშნავს ხსენებული *a priori* ოქსიმორული გამონათქვამი. ბერძნული წარმომავლობის ἄξιαπειον სიტყვა (როგორც ἄξιος «მახვილი» და μαρτύρι «სულელური» კომპონენტების შერწყმა) არის ერთმანეთის საწინააღმდეგო ცნებათა ხატოვანი კომბინაცია; წინააღმდეგობრივი ცნებების მახვილგონივრული შეხამება, პარადოქსი, სტილისტური ხერხი ან სტილისტური შეცდომა, საპირისპირო მნიშვნელობის სიტყვების კომბინაცია (ანუ შეუთავსებელი

სიტყვების შერწყმა). მაგალითად: ლევ ტოლსტოის პიესა «ცოცხალი ლეში», ნიკოლაი გოგოლის პოემა «მკვდარი სულელები», აიზეკ აზიმოვის რომანი «უსასრულობის დასასრული», ვსევოლოდ ვიშნევსკის პიესა «ოპტიმისტური ტრაგედია», მონაზონ სიმეონ ათონელის (წმინდა პანტელეიმონის რუსულ მონასტრიდან ათონზე) ქრისტიანული იგავი «ბრმა დარაჯი», ამერიკელი მწერალი ქალი სემილი ლოკხარტის რომანი «ნამდვილი ტყუილი», თელავის ოჯახური სასტუმრო «მცირე გრანდ ოტელი», «ფართოდ დახუჭული თვალებით მზერა» ...

მაგრამ ხელოვნური ინტელექტი გავლენას ახდენს თითოეულ ადამიანზე მის ყოველდღიურ საქმიანობაში და პირად ცხოვრებაშიც. მას ჯადოსნური მხარეც აქვს, ვინაიდან იგი შთამბეჭდავ შედეგებს იძლევა, ხოლო მისი მუშაობა და მექანიზმები ადამიანთა უმრავლესობისათვის სრულიად გაუგებარი რჩება.

ინფორმატიკა მიეკუთვნება ინფორმაციის რაციონალური დამუშავების სფეროს, კერძოდ, ავტომატური მანქანებით. ხელოვნური ინტელექტი კი არის ტექნოლოგია,

რომელსაც შეუძლია ადამიანის ტვინის მსგავსი შედეგების მიღება. ეს კომპიუტერული ინსტრუმენტი და იგი ასრულებს მოქმედებებს ან ხსნის ამოცანებს, რომლებიც უკანასკნელ დრომდე ცოცხალ არსებათა — ადამიანების ან ცხოველების — პრეროგატივა იყო.

ამერიკელი მეცნიერი ხელოვნური ინტელექტის დარგში, ტიურინგის პრემიის ლაურეატი მარვინ ლი მინსკი (Marvin Lee Minsky, 1927 - 2016), ერთ-ერთი იმ სწავლულთა შორის, რომლებმაც შემოიღეს და დაამკვიდრეს ტერმინი «*ხელოვნური ინტელექტი*», ასეთნაირად განსაზღვრავს ამ ცნებას: «*კომპიუტერული პროგრამების კონსტრუირება, რომლებიც დაკავებულია ადამიანების მიერ ამ მომენტისათვის უფრო დამაკმაყოფილებლად შესასრულებელი ამოცანებით, ვინაიდან ხსენებული ამოცანები მოითხოვს მაღალი დონის გონებრივ პროცესებს, როგორცაა პერცეპტუალური სწავლება (დასწავლა), მეხსიერების ორგანიზაცია და კრიტიკული დანასკვი (ინგლ. reasoning, inference)*». მარტივად რომ ვთქვათ, პერცეპტუალური (ლათ. perceptio — აღქმა) სწავლება არის სენსორულ სტიმულებზე რეაგირების —

*აღმის* — ცვლილებები, რომლებიც შექმნილია ამ სტიმულების განმეორებითი ზემოქმედებების შედეგად სპეციალური განმტკიცების (ინგლ. reinforcement), წახალისების გარეშე.

ინფორმატიკა — როგორც კომპიუტერული ინსტრუმენტი — დაფუძნებულია, კერძოდ, ალგორითმებზე, ე. ი. მათემატიკური ფორმულებისა და სტატისტიკური დამუშავების მიმდევრობებზე. იგი მუშაობს «შემავალ სიდიდეებთან» (საწყის მონაცემებთან) და აღწევს «გამომავალ სიდიდეებს» (შედეგებს) სხვადასხვა ბიჯის, ეტაპის შესრულებით, რომლებიც ითხოვს გამოთვლებს, ლოგიკურ ოპერაციებს, შედარებებს ან ანალოგებს. მაგრამ, ხელოვნური ინტელექტი მიეკუთვნება ტექნოლოგიურ მოწყობილობებს, რომლებიც მიზნად ისახავს ბუნებრივი ინტელექტის იმიტაციას და საბოლოო ჯამში მის ჩანაცვლებას. იგი ცდილობს ასახოს ადამიანისა და ცხოველის უნარი აღიქვას, განასხვაოს, გაიგოს, ისწავლოს, იმსჯელოს, გამოთვალოს, დაიმახსოვროს, შეადაროს, აირჩიოს და ა.შ. თუმცა ხელოვნური ინტელექტი ბიოლოგიური ინტელექტის ფუნქციონირების იმიტირებას კი არ

ცდილობს, არამედ მხოლოდ შედეგების, რომლებსაც ეს უკანასკნელი აღწევს — კავშირი მათ შორის დაახლოებით ისეთივეა, როგორც თვითმფრინავსა და ფრინველთა ფრენას ან ავტომობილსა და რბენას შორის. ამ სფეროს მკვლევართა მიზანი არ არის და არც ყოფილა ტექნოლოგიური ტვინის შექმნა, რომელიც მაქსიმალურად უახლოვდება ბუნებრივ ტვინს, ადამიანის გონებას და აზროვნებას; ლაპარაკია მხოლოდ და მხოლოდ კომპიუტერული ინსტრუმენტების შექმნაზე, რომლებსაც შეუძლია ყველაზე უფრო რელევანტური — მომხმარებლის მოთხოვნების შესაბამისი — შედეგების დადება, ძირითადად, მონაცემთა სტატისტიკური ანალიზის საშუალებით.

სიტყვა «ინტელექტი» ნაწარმოებია ლათინური სიტყვიდან, რომელიც მიშნავს გაგებას, რისამე გარჩევას, განსხვავების დანახვას, ხოლო ამ სიტყვის პრეფიქსი ითარგმნება როგორც რისამე «შორის». ესე იგი, ინტელექტუალურია ის, რასაც აქვს უნარი გააკეთოს არჩევანი სხვადასხვა ალტერნატივას შორის და ამისათვის შეადაროს

საგნები ერთმანეთს, რათა უკეთ შეიგნოს ისინი. ადამიანებს და ცხოველებს, რა თქმა უნდა, აქვს ეს უნარი. ამრიგად, ხელოვნური ინტელექტის «ინტელექტუალური» ასპექტი შეესაბამება ადამიანის ან ცხოველის ქცევის იმიტირების მიზანსა და უნარს, ხოლო «ხელოვნური» ასპექტი მიეკუთვნება კომპიუტერებისა და სხვა ელექტრონული და ტექნოლოგიური პროცესების გამოყენებას. მაშასადამე, ხელოვნური ინტელექტი იღებს სესხად იდეებს ინფორმატიკის, ლოგიკის, მათემატიკის, ელექტრონიკის, კოგნიტიური თეორიების, კომუნიკაციური მეცნიერების, ასევე ნეირობიოლოგიისა და ეთოლოგიისგანაც კი — მეცნიერებიდან, რომელიც შეისწავლის ცხოველთა, ძირითადად, ბუნებრივ პირობებში ქცევის ბიოლოგიურ საფუძვლებს. იგი ახორციელებს მათ შეთავსებას ღრმა სწავლების მეთოდებთან, რომლებიც ვითარდება ურთიერთდაკავშირებული ნეირონული ქსელების (ანუ *კონექციონიზმის*) ბაზაზე.

მაგრამ, მიუხედავად იმისა, რომ ხელოვნურ ინტელექტს შეუძლია ცოდნათა თავისი სისტემის ან მონაცემთა აღქ-



მის გადაწყობა ახალი მიღებული ინფორმაციის შესაბამისად, არც ერთი კომპიუტერი, ალგორითმი ან ნეირონული ქსელი არ ფიქრობს და არ არის გონიერი ამ სიტყვის მკაცრი გაგებით — მისი სწავლებისა და გადაწყვეტილების მიღების უნარი დაპროგრამებულია ან ავტომატიზებულია, რაც ინტელექტის მხოლოდ ილუზიას ქმნის, უკეთეს შემთხვევაში — ძალიან სუსტი ინტელექტის, ვინაიდან იგი პრაქტიკულად მექანიკურია. ხელოვნურ ინტელექტს არ შეუძლია საკუთარი ამოცანების ავტონომიურად ამოხსნა ან ისეთი ახალი წარმოდგენების შექმნა სამყაროს შესახებ, რომლებითაც იგი იხელმძღვანელებს თავის მსჯელობებში; თუმცა არავითარი მსჯელობა მისთვის ფაქტობრივად არც კი არსებობს, ვინაიდან იგი მხოლოდ რეფლექსი და ავტომატიზმია. სწორედ ამიტომ ტერმინი «*ხელოვნური ინტელექტი*» ძალიან გადაჭარბებულია.

სინამდვილეში საქმე ეხება არა ინტელექტს, არამედ მხოლოდ გამოთვლებს (ვადიაროთ ესეც, *ნასწავლ — d'é-jà vu* — *გამოანგარიშებებს*), სტატისტიკას და კომპიუტერულ ოპერაციებს.

ამიტომ «ხელოვნურ ინტელექტზე» ლაპარაკი ენობრივად ამ კონსტრუქციის გაუმართლებელი გამოყენებაა. «ხელოვნური ინტელექტი» ნაწარმოებია ინგლისური სიტყვებით *«artificial intelligence»*. უილიამ შექსპირის (William Shakespeare, 1564 - 1616) ენაში ეს გამონათქვამი ეხება მონაცემთა მარტივ მართვას, ინფორმაციის დამუშავებას და არა ქართულ ან, ვთქვათ, ფრანგულ ენაში გაგებული ინტელექტის ხელოვნურ ნაირსახეობას. ხსენებულ ენებში ტერმინი «ხელოვნური ინტელექტი» უფრო შეიძლება ითარგმნოს როგორც *«ხელოვნური საზრია-ნობა»*, რაც ნიშნავს *საღ აზრს, ემკათიას* (ადამიანის შესაძლებლობას თანაუგრძნოს სხვებს და გაიგოს მათი ფსიქოლოგიური მდგომარეობა), *დასაბუთებულ ნებას, საკუთარი ინტელექტის გაგებისა და შეცნობის უნარს*. მოკლედ რომ ვთქვათ, ინგლისურ ენაში ტერმინის «ხელოვნური ინტელექტი» ხმარება ამ ახალი ტექნოლოგიების აღნიშვნის მიზნით გამართლებულია და არ იქნება ადამიანთა შეცდომაში შეყვანა. მაგრამ ქართულ ენაში, როგორც, სხვათა შორის, ფრანგულშიც, ხელოვნურ ინტელექტზე ლაპარაკი საკმაოდ უხერხულია.

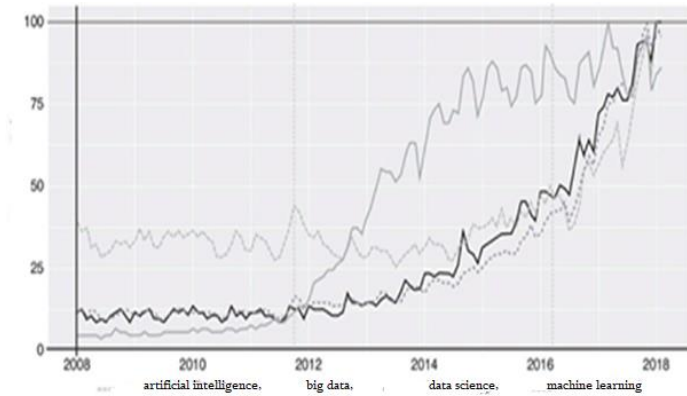
ამჟამად, რობოტი არ მსჯელობს, არაფრის სჯერა, თითქმის არ იღებს საკუთარი ნებით გადაწყვეტილებას, არ აქვს განზრახვა ან ყურადღება, არ შეუძლია ახალი მიზნების და ამოცანების დასახვა, არ აგებს რეალობას აზროვნებაში, არ შეუძლია კონცეფციის შემუშავება ან იდეის წამოყენება, არ გააჩნია კულტურა და მით უმეტეს ზოგადი, არ აქვს თვითცნობიერება და გადარჩენის ინსტინქტი — ინტელექტის ყველა მთავარი ნიშანი. ადამიანს ახასიათებს მახვილგონიერება, თავისუფლება და უნარი გაიხსნას უცნობთან. რობოტი სრულიად განსხვავებულია მისგან. ამიტომ ლაპარაკი «ხელოვნურ ინტელექტზე» — ეს თითქმის იგივეა, რაც ლაპარაკი «ხელოვნურ კაცობრიობაზე». «ხელოვნური ინტელექტი» არის *ანთროპომორფიზმი* (ადამიანურ თვისებათა გადატანა ბუნებრივ მოვლენებზე, მცენარეებსა და ცხოველებზე). ჩვენ ვაადამიანებთ და ვაცოცხლებთ იმას, რაც ჯერ კიდევ მხოლოდ ანიმირებული ობიექტია, ავტომატი და სუპერ კომპიუტერი. ტიურინგის პრემიის ლაურეატი ფრანგი მკვლევარი იან ლეკუნი (ფრანგ. Yann André Le Cun, 1960) ამის მიუხედავად, მაინც

თვლის, რომ ხელოვნური ინტელექტი მალე შეძლებს გრძნობებისა და ემოციების გამოხატვას. ამავდროულად იგი მიანიშნებს, რომ ეს მოხდება მხოლოდ იმიტომ, რომ ხელოვნური ინტელექტი შეძლებს სიხარულის, შიშის ან სევდის შეფასებას რაოდენობრივად. მაგრამ ასეთნაირად *გაანგარიშებული* სიხარული, შიში ან სევდა, რომლებსაც არ შეიგრძნობენ და არ განიცდიან, შეიძლება ნამდვილ ემოციებად მივიჩნიოთ? იქნებ, ჯერ უკეთესი იყოს ლაპარაკი «ავტომატურ სწავლებაზე», «სტატისტიკურ» ან «პროგნოსტიკურ» ინსტრუმენტებზე, «ავტონომიურ» ან «ადაპტირებად» აგენტებზე? დღეს ხელოვნური ინტელექტი კომპიუტერულ ტომოგრაფიას ახორციელებს უკეთ, ვიდრე რენტგენოლოგი, კიბოს დიაგნოზს სვამს უკეთ, ვიდრე ონკოლოგი, ავტომობილს მართავს უკეთ, ვიდრე მძღოლი, მაგრამ მან არ იცის, რა არის რენტგენის გამოსხივება, რა არის ავთვისებიანი სიმსივნე, რა არის ავტომობილი ან რას ნიშნავს მისი ტარება.

ვინაიდან ტერმინი «ხელოვნური ინტელექტი» (ანუ ინგლისური აბრევიატურით «AI» ფრაზიდან *Artificial*

*Intelligence*) ფართოდ გამოიყენება როგორც საჯარო დებატებში, ასევე სპეციალიზებულ დისკუსიებში და სამეცნიერო კვლევათა ახალი სფეროს დასახელებაშიც კი — ძნელია ამ ნარკვევის როგორღაც სხვანაირად დასათარება, ვინაიდან იგი (რკვევის საგანი) გაუგებარი დარჩება. უფრო მეტიც, რაკი ეს ყოვლისმომცველი ტერმინია, რომელიც ნიშნავს განსხვავებულ და მრავალფეროვან ტექნოლოგიებს, უფრო ლოგიკური იქნებოდა მის ნაცვლად გველაპარაკა «ხელოვნურ ინტელექტებზე» მრავლობითი რიცხვის გამოყენებით. მაგრამ ხმარების გამოცდილება უპირატესობას მხოლოდით რიცხვს აძლევს, როცა საუბარია «ხელოვნური ინტელექტის» ზოგად ცნებაზე...

მოდა «ხელოვნურ ინტელექტსა» და მის ინგლისურ «AI» აბრევიატურაზე არც ისე შორეულ წარსულში გაჩნდა, რაზეც მეტყველებს ყველაზე ცნობილ საძიებო სისტემაში — ინტერნეტში — შესულ შეკითხვათა სტატისტიკა (ნახ. 5).



ნახ. 5 ძებნა ინტერნეტში ხელოვნურ ინტელექტთან დაკავშირებულ თემათა მიხედვით

ეს მოდა იმდენად ძლიერია, რომ, თუ გვინდა ვისაუბროთ ხელოვნური ინტელექტის შესახებ, ყოველთვის შეგვიძლია ვიხმაროთ მხოლოდ მისი აბრევიატურა.

მე-5 ნახატი მოცემული გვაქვს ლინო გალიანას (Lino Galiana) ნაშრომის [7] საფუძველზე, სადაც წყაროს წარმოადგენს Google Trends საიტის მონაცემები. ისინი ნორმალიზებულია. ყოველი შეკითხვისათვის ნორმალიზაცია ხდება პროცენტებში დროის გარკვეულ მონაკვეთზე დანაკვირვებ შეკითხვათა მაქსიმალური რიცხვიდან.

Google Trends — ეს Google კომპანიის ვებ-საიტია, რომელიც ანალიზებს საუკეთესო საძიებო შეკითხვათა პოპულარობას Google-ის ძიებაში სხვადასხვა რეგიონსა და სხვადასხვა ენაზე.



ნახ. 6 «შესანიშნავი ოთხეული»

1956 წელს ამერიკელ მათემატიკოსთა «შესანიშნავი ოთხეული» (ნახ. 6) — ჯონ მაკარტი (John McCarthy, 1927—2011), მარვინ მინსკი (Marvin Lee Minsky, 1927—2016), კლოდ შენონი (Claude Elwood Shannon, 1916—2001) და ნატანიელ როჩესტერი (Nathaniel Rochester, 1919—2001)

— მოევლინა მსოფლიოს მეცნიერთ «*მოაზროვნე მანქანების*» თემისადმი მიძღვნილი სემინარის ორგანიზატორად კერძო კვლევით უნივერსიტეტში — დარტმუთის კოლეჯში (Dartmouth College) ქალაქ ჰანოვერში (Hanover), შტატი ნიუ-ჰემპშირი (New Hampshire).

სემინარზე პირველად იყო შემოთავაზებული ტერმინი «ხელოვნური ინტელექტი».

ბრიტანელი მათემატიკოსი ალან ტიურინგი (Alan Mathison Turing, 1912—1954) ასევე ითვლება ხელოვნური ინტელექტის *იდეის ავტორად*, თუმცა არა თავად ამ ტერმინის (ნახ. 7).



ნახ. 7 ალან ტიურინგი — ისტორიაში ერთ-ერთი 100 უდიდეს ბრიტანელს შორის



უინსტონ ჩერჩილი (Winston Churchill, 1874-1965) თვლიდა, რომ ტიურინგმა ყველაზე დიდი წვლილი შეიტანა მოკავშირეთა გამარჯვებაში ნაცისტური გერმანიისა და მისი პარტნიორების წინააღმდეგ ომში.

ხოლო მეცნიერის ერთ-ერთ კოლეგას ბლეტჩლი-პარკიდან (ნახ. 8) — დეშიფრაციის ლეგენდარულ ბრიტანულ ცენტრიდან მეორე მსოფლიო ომის დროს — ნათქვამი აქვს, რომ *«ბლეტჩლი-პარკს ესაჭიროებოდა განსაკუთრებული ტალანტი, განსაკუთრებული გენიალობა და ტიურინგის გენიალობა იყო სწორედ ასეთი»*.



ნახ. 8 დეშიფრაციის ბრიტანული ცენტრის რეზიდენცია

Bletchley Park

პირად ცხოვრებაში ალან ტიურინგი იყო ადამიანი, რომელსაც წაართვეს სიყვარულის უფლება.

1952 წლის იანვარში იგი გაქურდეს და მისგან პოლიციაში სათანადო განცხადება შევიდა. პოლიციამ სულ რამდენიმე დღეში მიაგნო ქურდს. აღმოჩნდა, რომ ეს იყო ვინმე 19 წლის მუშა არნოლდ მიურეი (Arnold Murray, 1932-1990). დაკითხვაზე უკანასკნელმა აღიარა, რომ მან, მართალია, წაიღო ფული, მაგრამ ეს იყო არა ქურდობა, არამედ სარგო მის მიერ გაწეული სექსუალური მომსახურებისათვის. პოსტვიქტორიანულ ბრიტანეთში ჰომოსექსუალიზმი ისჯებოდა. ალან ტიურინგი ან უნდა წასულიყო ციხეში, ან დათანხმებულიყო ალტერნატივაზე — იძულებით ჰორმონულ მკურნალობაზე ლიბიდოს გასაქრობად, ფაქტობრივად, ქიმიურ კასტრაციაზე. ტიურინგმა მეორე — ქიმიური კასტრაცია — აირჩია. განა ალან ტიურინგის სასჯელი მეოცე საუკუნის ბრიტანეთში ბევრად უფრო ჰუმანურია, ვიდრე ფარსმან სპარსის სასჯელი მეთერთმეტე საუკუნის საქართველოში?!

ერთი წლის განმავლობაში ტიურინგს მკურნალობდნენ ჰორმონული პრეპარატის — *სტილბესტროლის* — ინიექციებით. ამან გამოიწვია იმპოტენცია, ლიბიდოს გაქრობა და გინეკომასტია (გულმკერდის ჯირყვლების გამკვრივება და გადიდება, მკერდის მიერ გიგანტური სილიკონის მკერდის ფორმის მიღება). დღეს ამ პრეპარატით მკურნალობენ ქალებს, რომლებსაც აქვთ სუსუნატით (ტრიპერით, ჰონორეით) გამოწვეული ვაგინიტი. 1954 წლის 8 ივნისს ალან ტიურინგი გარდაცვლილი იპოვეს თავის ბინაში. გამოძიებამ დაადგინა, რომ მეცნიერმა სიცოცხლე თვითმკვლელობით დაასრულა.

მამათმავლობა ინგლისში და უელსში რჩებოდა სისხლის სამართლის დანაშაულად კიდევ 13 წელი, 1967 წლამდე.

2009 წელს ბრიტანეთის პრემიერ-მინისტრმა გორდონ ბრაუნმა (James Gordon Brown, 1951) საჯაროდ მოიხადა ბოდში დევნისათვის, რომელიც განიცადა ალან ტიურინგმა. თავისი გამოსვლა პრემიერმა დაასრულა სიტყვებით: *«ბრიტანეთის მთავრობის და ყველას სახელით, ვინც ცხოვრობს თავისუფლების პირობებში*

*ალანის წვლილის წყალობით, ვამბობ: გვაპატიე, შენ გაცილებით უკეთეს ბედს იმსახურებდი».*

2013 წელს დიდი ბრიტანეთის დედოფალმა ელისაბედ მეორემ ოფიციალურად შეიწყალა ტიურინგი «ურცხვ საქციელში» ბრალისათვის, რომელიც მას ედებოდა. ელისაბედ მეორის თითქმის სამოცწლიანი მმართველობის ისტორიაში, ეს შეწყალების მხოლოდ მეოთხე შემთხვევად დარჩა.

საქართველოში კარგად არის ცნობილი ალან ტიურინგის ცხოვრების ამსახველი 2014 წლის ისტორიული დრამა «იმიტაციის თამაში», რომელშიც მთავარი გმირის როლს სახელგანთქმული (განსაკუთრებით, შერლოკ ჰოლმსის განსახიერებით) ბენედიქტ ტიმოტი კარლტონ კამბერბეტი (Benedict Timothy Carlton Cumberbatch, 1976) ასრულებს. რაც შეეხება ტიურინგის მიერ სპეციალური კონკურსის შედეგად აღმიჩენილი უნიჭიერესი მათემატიკოსის და შემდეგ ახლო მეგობრისა და საქმის ერთგული კოლეგის ჯოან ელისაბედ კლარკს (Joan Elisabeth Lowther Clarke, 1917-1996), მის როლში მაცურებელს ბრიტანული კინოს შეუდარებელი

ვარსკვლავი კირა ქრისტინა ნაიტლი (Keira Christina Knightley; 1985) ხიბლავს (ნახ. 9).



ნახ. 9 კირა ნაიტლი და ბენედიქტ კამბერბეტი

აღსანიშნავია, რომ ალან ტიურინგი იმდენად აფასებდა ჯოან კლარკს და წარმოუდგენლად მიაჩნდა მისი წასვლა დემიფრატორთა გუნდიდან, რომ ამის თავიდან ასაცილებლად 1941 წელს წინადადებაც კი გაუკეთა კოლეგას გამოჰყოლოდა ცოლად, მაგრამ ნიშნობა ხანმოკლე აღმოჩნდა. თავისი ჰომოსექსუალობის აღიარე-

ბისას საცოლესთან, უკანასკნელი ამ აღიარებას სრულად გულგრილად და აუღელვებლად შეხვდა, თუმცა მალე ორივემ ისურვა უარი ეთქვა დაქორწინებაზე.

ბედის ირონია ის არის, რომ ერთისა და იმავე Murray გვარის მატერებელმა ორმა მამაკაცმა ერთსა და იმავე 1952 წელს ავის მომასწავებელი, თვალბედიითი ყორნის როლი ითამაშა ალან ტიურინგის ცხოვრებაში: არნოლდ მიურეიმ (Arnold Murray) იგი ქიმიურ კასტრაციამდე მიიყვანა, ხოლო ჯონ მიურეიმ (John Kenneth Ronald Murray) მას ყველაზე ახლო მეგობარი და ფასდაუდებელი თანამშრომელი ჯოან კლარკი (ნახ. 10) წაართვა და ცოლად შოტლანდიაში წაიყვანა.



ნახ. 10 ჯოან კლარკი (გათხოვების შემდეგ სიკვდილამდე, 1952-1996 წლებში, Joan Clarke Murray)

ხელოვნური ინტელექტის პრობლემით ტიურინგი და-  
ინტერესდა ჯერ კიდევ 1950 წელს, როცა მუშაობდა  
«იმიტაციის თამაშზე» (ანუ «ტიურინგის ტესტზე»).

ტიურინგის ტესტი — ემპირიული ტესტია, რომლის  
იდეა წამოყენებული იყო ალან ტიურინგის მიერ სტატი-  
აში «გამოთვლითი მანქანები და გონება», («Computing  
Machinery and Intelligence»). იგი გამოქვეყნდა 1950 წელს  
ბრიტანულ ანალიტიკურ ფილოსოფიურ ჟურნალში  
*Mind* (ინგლ. «ტვინი, გონება, გონი»). ტიურინგმა მიზ-  
ნად დაისახა იმის გარკვევა, შეუძლია თუ არა მანქანას  
იაზროვნოს.

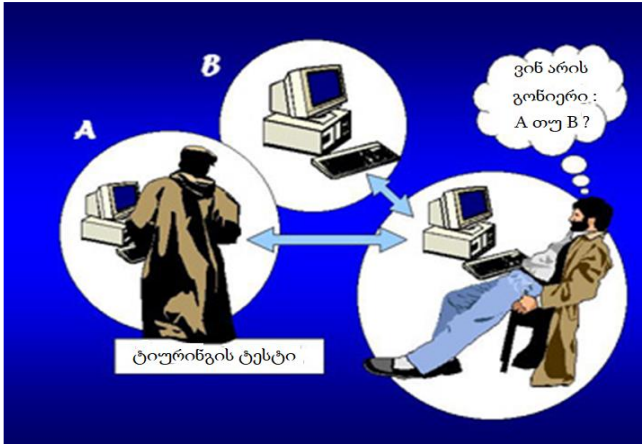
ამ ტესტის სტანდარტული ინტერპრეტაცია შეიძლება  
შემდეგნაირად ჩამოყალიბდეს: *«ადამიანი ურთიერთ-  
ქმედებს ერთ კომპიუტერთან და ერთ ადამიანთან.  
კითხვებზე გაცემული პასუხების საფუძველზე, მან უნ-  
და დაადგინოს, ვისთან ლაპარაკობს: ადამიანთან თუ  
კომპიუტერულ პროგრამასთან. კომპიუტერული პრო-  
გრამის ამოცანაა — შეიყვამოს ადამიანი შეცდომაში და  
აიძულოს იგი გააკეთოს მცდარი არჩევანი»*.

ტესტის არც ერთი მონაწილე ერთმანეთს არ ხედავს. თუ მსაჯს არ შეუძლია გარკვეულად თქვას, ვინ არის მოსაუბრეთა შორის ადამიანი, მაშინ ითვლება, რომ მანქანამ ტესტი გაიარა. იმისათვის, რომ მოხდეს სწორედ მანქანის ინტელექტის ტესტირება და არა მისი შესაძლებლობის გაარჩიოს ზეპირმეტყველება, საუბარი მიმდინარეობს «*მხოლოდ ტექსტის*» რეჟიმში, ვთქვათ, დამხმარე («*შუამავალი*») კომპიუტერის კლავიატურის ან ეკრანის საშუალებით. მიმოწერა უნდა ხდებოდეს დროის კონტროლირებად შუალედებში, რათა მსაჯმა ვერ შეძლოს დასკვნის გამოტანა პასუხების სიჩქარეებიდან გამომდინარე. ტიურინგის დროს კომპიუტერები ადამიანზე ნელა რეაგირებდა. ახლა ეს წესი ასევე აუცილებელია, მაგრამ უკვე იმიტომ, რომ კომპიუტერი ადამიანზე სწრაფად რეაგირებს.

ერთი სიტყვით, ეს ტესტი იმაში მდგომარეობს, რომ ადამიანს (მსაჯს) აყენებენ ფარულ სიტყვიერ დაპირისპირებაში კომპიუტერთან და სხვა ადამიანთან (ნახ. 11);



თუ პირველ ადამიანს (მსაჯს) არ შეუძლია თქმა, მის თანამოსაუბრეთა შორის რომელია კომპიუტერი, მაშინ კომპიუტერმა გაიარა ტესტი [8].



ნახ. 11 ტიურინგის ტესტი (1950)

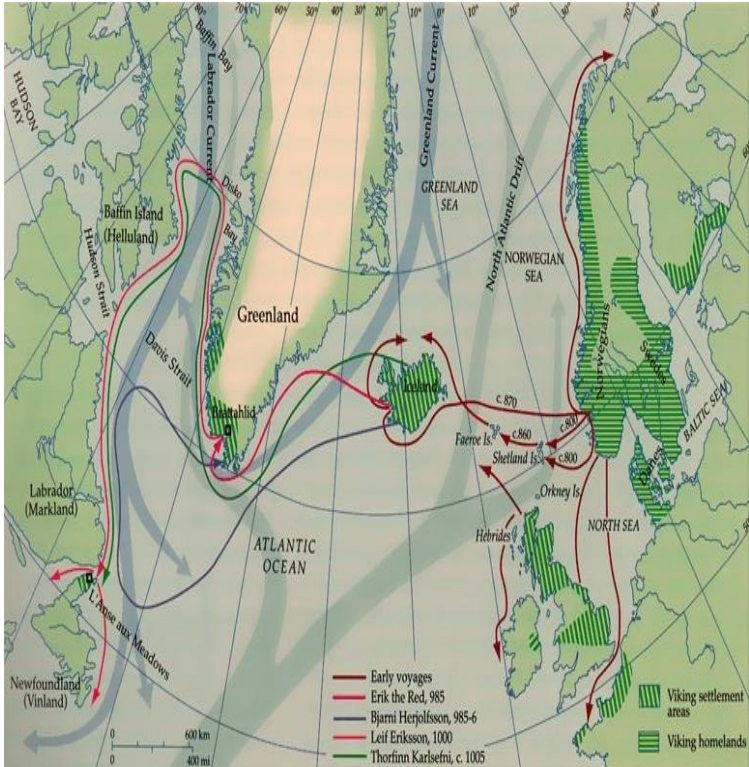
ტიურინგი ვარაუდობდა, რომ მანქანები ბოლოს და ბოლოს შეძლებს ტესტის გავლას ; ფაქტობრივად იგი მოელოდა, რომ 2000 წლისათვის მანქანები  $10^9$  ბიტ მოცულობის მეხსიერებით მოახერხებს მსაჯთა 30%-ის შეყვანას შეცდომაში ხუთწუთიანი ტესტის შედეგების მიხედვით.

გაჩნდა განათლებისა და მეცნიერული კვლევების ახალი სფერო, რომლის მიზანია — ოდესღაც ასე მიჩნეული და დღემდე მოგონილ არარეალობად დარჩენილი — სწავლის შემძლე მანქანების შექმნა. სახელდობრ, ამ მანქანამ უნდა ისწავლოს, განსაკუთრებით, საკუთარ შეცდომებზე, გაუგოს გარემოს და სამყაროს მთლიანობაში, შეეწყოს ცვლილებებს, მოახერხოს მომავლის წინასწარმეტყველება და პროგნოზირება, ასევე მოქმედება ადამიანის ცხოვრების პირობათა გასაუმჯობესებლად. რაც უფრო სრულყოფილად იქნება მიღწეული ეს თავდაპირველი მიზანი, მით უფრო ნაკლებ იმედგაცრუებას განიცდის ადამიანი ხელოვნური «ინტელექტის» მიმართ. მაშინ შესაძლებელი გახდება სრული უფლებით შევადართ იგი ბიოლოგიურ ინტელექტს.

მაგრამ ჯერჯერობით ის, რასაც «ხელოვნურ ინტელექტად» ნათლავენ, რჩება ადამიანის ინტელექტის პროდუქტად, მის ქმნილებად, შემოქმედებად და ასახვად, თუმცა მთლიანობაში ძალიან ჩამორჩება მას. რეალობად რომ იქცეს, ხელოვნურმა ინტელექტმა უნდა შეიძინოს ავტონომია და ნების თავისუფლება, გამოეყოს თავის

მშობელ ადამიანებს. მას მოუწევს გახდეს საყოველთაო და ზოგადი ხელოვნური ინტელექტი, რომელსაც აქვს უნარი გაიგოს, შეიცნოს სამყაროს ფუნქციონირება მთელი მისი სიღრმით და სისრულით მარტივი დაკვირვების გზით, როგორც ამას ბავშვი აკეთებს თავისი ცხოვრების პირველ წლებში. მაგრამ ამის მიღწევა ადვილი არ იქნება, თუმცა მანქანური სწავლების გამოყენება მის მიმართაც კი დაიწყოს, რომ აიძულონ კომპიუტერი ააგოს ნეირონული ქსელები. ასე რომ ხელოვნური ინტელექტები ოდესმე შეძლებს სხვა ხელოვნური ინტელექტების შექმნას. ძნელი წარმოსადგენია, რომ ხელოვნური ინტელექტი მიუერთდება ბუნებრივ ინტელექტს და გაუსწრებს მას. ასეთი რამ, ალბათ, შეიძლება მოხდეს მხოლოდ ძალიან სპეციფიკურ ამოცანებში, რომლებიც, სავარაუდოდ, დაფუძნებული იქნება მონაცემთა სტატისტიკურ დამუშავებაზე. თუმცა ამოხსნის მრავალი პროცესის წარმართვა შესაძლებელია სხვადასხვა გზით და არ გულისხმობს სტატისტიკური მიდგომების სავალდებულო გამოყენებას ან გაბატონებული სტატისტიკური ტენდენციების კალაპოტში დარჩენას.

როცა მკვლევარებმა დარტმუთის კოლეჯიდან გამოიგონეს ტერმინი «ხელოვნური ინტელექტი», ძირითადად, ეს იყო მარკეტინგული სვლა, თითქმის მიზანმიმართული სიცრუე: გამოიყენეს სენსაციური ტყუილი, რათა შემდეგ მიეღოთ კვლევების დაფინანსება ამერიკის შეერთებული შტატების მთავრობისგან. ეს რეალობის შეგნებული დამახინჯებაა, რომელიც გვაგონებს რეალობის დამახინჯებას *წითური ეირიკის* მიერ. ისლანდიური «საგა წითურ ეირიკზე» (*Eiríks saga rauða*) გადმოგვცემს, რომ ეირიკმა წაიყვანა თავისი ხალხი გრენლანდიაში და აიძულა ისინი ერწმუნათ, რომ ეს «*მწვანე ქვეყანაა*» (რადგან სწორედ ასე ითარგმნება სიტყვა «გრენლანდია» — «groenland»), მაშინ როცა სინამდვილეში ეს იყო თოვლითა და ყინულით დაფარული *თეთრი ქვეყანა*. მაგრამ სახელწოდება «გრენლანდია» დარჩა და ხალხი ამ სახელს შეაჩვიეს. ისლანდიური საგა ვიკინგების მოგზაურობებზე ეძღვნება ამერიკის აღმოჩენას კოლუმბამდე დიდი ხნით ადრე (ნახ. 12).



ნახ. 12 ვიკინგების მოგზაურობა გრენლანდიაში და ე. წ. ვინლანდში (ამერიკაში).

ამ ნახაზზე წარმოდგენილი რუკა რედაქტირებულია მაიკლ დუმას (Michael Douma) მიერ, რომელიც არის ჯორჯტაუნის უნივერსიტეტის ბიზნესის მაკდონოს სკოლის მკვლევარი ასისტენტ-პროფესორი, სადაც იგი

ბაზრისა და ეთიკის შემსწავლელი ჯორჯტაუნის ინსტიტუტის დირექტორია (Assistant Research Professor in the McDonough School of Business at Georgetown University, where he is the Director of the Georgetown Institute for the Study of Markets and Ethics). ამავდროულად გატაცებულია ამერიკის კონტინენტის ისტორიით და პროფესიულად იკვლევს მისი აღმოჩენის პრობლემებსაც.

აქედან კარგად ჩანს, რომ წითური ეირიკის (Erik the Red) მოგზაურობა და ამერიკის აღმოჩენა თარიღდება 985 წლით, მაშინ როცა ქრისტეფორე კოლუმბის (ლათ. Christophorus Columbus) — ესპანეთის გვირგვინის სამსახურში მყოფი გენუელი მეზღვაურის — მიერ ამერიკის აღმოჩენა ოფიციალურად თარიღდება 1492 წლის 12 ოქტომბრით.

ინტელექტის კონცეფციას ონტოლოგიურად (ე. ი. ყოფიერების ანუ არსებობის, არსის, სტრუქტურის თვალსაზრისით) აქვს ბიოლოგიური საფუძველი და «ხელოვნურ» ინტელექტი, განსაზღვრებითაც კი, არ შეიძლება იყოს გონიერი. თუმცა ეს ხელს არ უშლის ხელოვნურ ინტელექტს გაუწიოს ადამიანს დიდი სამსახური და

ერთდროულად შეუქმნას ბევრი გამოწვევა ადამიანურ ინტელექტს, პირველ რიგში სწორედ მის ინტელექტს, ე. ი. სამყაროს მისებურ გაგებას, აღქმას, ახსნასა და კრიტიკულ ანალიზს.

რაც შეეხება «შესანიშნავი ოთხეულის» მიზანმიმართულ მარკეტინგულ სიცრუეს, მათი მიზანი შეიძლება ამართლებდეს კიდევ ამ ქცევას. და ეს იეზუიტთა ორდენის დამაარსებლისა და მისი პირველი გენერლის *იგნატიუს (იგნასიო) ლოიოლას* (ლათ. Ignatius de Loyola, 1491–1556) კრედო კი არ არის, არამედ *შალვა ნუცუბიძის* (1888-1969) აღმოსავლური რენესანსის (*რომელიც ორასი წლით მაინც, თუ მეტით არა, უსწრებდა ევროპულს*) უდიდესი მნათობის *შოთა რუსთველის* სამოქმედო პრინციპი: მისი გმირები, ცრუობენ, მრუშობენ, ჰკლავენ, დალატობენ, მაგრამ მიმზიდველები რჩებიან ჩვენთვის, რადგან მათი მიზანი უკეთილშობილესია — უზომო სიყვარული.

### III. ჩინური ოთახი

ჩინური ოთახი (ინგლ. chinese room) — პირველად 1980 წელს ამერიკელი ფილოსოფოსის ჯონ როჯერს სირლის (John Rogers Searle, 1932) მიერ გამოქვეყნებული სააზროვნო ექსპერიმენტია ცნობიერებისა და ასევე ხელოვნური ინტელექტის ფილოსოფიაში. ექსპერიმენტის მიზანი, მდგომარეობს *იმ მტკიცების უარყოფაში*, რომ ციფრულ მანქანას — მისი გარკვეული ხერხით დაპროგრამების შედეგად შექმნილი «ხელოვნური ინტელექტი» — აქვს უნარი იქონიოს ცნობიერება იმავე აზრით, რომლითაც იგი ადამიანს გააჩნია. სხვა სიტყვებით, მიზანია ეგრეთ წოდებული «ძლიერი» ხელოვნური ინტელექტის ჰიპოთეზის უარყოფა და ტიურინგის (Alan Mathison Turing, 1912-1954) ტესტის კრიტიკა.

წარმოვიდგინოთ, რომ მეცნიერებმა გამოიგონეს ზემდლავრი ჩიპი, რომლის წყალობით მანქანებმა შეიძინა არნახულ შესაძლებლობათა ხელოვნური ინტელექტი.

არც თუ შორეულ მომავალში ისინი უკონტროლო გახდება, მანქანებმა გაანადგურა კაცობრიობა და დაიპყრო დედამიწა.



შესაძლოა მოვლენათა განვითარების მეორე ვარიანტი. მანქანებმა, რომლებმაც თავიანთ განვითარებაში ადამიანს ათასჯერ გაუსწრო, მონებად აქცია შემქმნელები და იყენებს მათ შინაურ ცხოველებად...

ალბათ, სიუჟეტი ყველასათვის კარგად არის ცნობილი, მის საფუძველზე დაწერილია უამრავი ფანტასტიკური რომანი და გადაღებულია ფილმი.

*მაგრამ შესაძლებელია სინამდვილეში ხელოვნურმა გონებამ გაუსწროს ადამიანისას ან თუნდაც მიუახლოვდეს მას?*

დისკუსია ამ თემაზე არ ქრება კულტურაში «კიბერნეტიკული ხანის» დაწყების დღიდან. გონიერ რობოტებს უძღვნიდნენ თავიანთ ნაწარმოებებს ამერიკელი ფანტასტები – ავილოთ თუნდაც ცნობილი მეცნიერი-ბიოქიმიკოსი აიზეკ აზიმოვი (Isaac Asimov, 1919 — 1992) მისი «რობოტიქსის სამი კანონით» («Three Laws of Robotics», 1991). სხვათა შორის იმ წლებში ე.წ. «ელექტრონული გამომთვლელი მანქანების» რაოდენობა სულ რამდენიმე ცალს შეადგენდა და მათი ზომები დიდი ოთახის ზო-

მებს უდრიდა, ისინი განლაგებული იყო მსხვილ სამეცნიერო-კვლევით ინსტიტუტებში და ცოტა რამ თუ ეხერხებოდა ინტელექტუალური თვალსაზრისით. მაგრამ მაშინაც კი მოუსვენარი პოლონელი ფილოსოფოსი, სატირიკოსი, მწერალი და ფუტუროლოგი სტანისლავ ლემი (Stanislaw Herman Lem, 1921 — 2006) დაჟინებით ამტკიცებდა, რომ მანქანის ინტელექტი და ადამიანის გონება — «ორი დიდი განსხვავებული რამ არის».

მსგავსი აზრისა იყო ამერიკელი ფილოსოფოსი ჯონ როჯერს სირლი (John Rogers Searle; 1937), რომელმაც 1980 წელს თვალსაჩინოდ ახსნა განსხვავება ინფორმაციის მანქანურ დამუშავებასა და ადამიანის აზროვნებას შორის. თავის წარმოსახვით ექსპერიმენტს მან «ჩინური ოთახი» უწოდა. წარმოიდგინეთ, რომ ჩაგკეტეს ოთახში, სადაც წარწერებია თქვენთვის უცნობ ენაზე — ასეთ ენად სირლიმ შეარჩია ჩინური («ჩინური დამწერლობა» ხომ უძველესი დროიდან ასოცირებულია რაღაც გაუგებართან და მიუწვდომელთან), მაგრამ ეს შეიძლება იყოს ნებისმიერი სხვა ენა, თუნდაც უცხო პლანეტის ან ამერიკის ინდიელთა ჯგუფის *ნავახოს* (ინგლ. *Navajo*) ტომის

ენა, რომელსაც ამერიკის შეერთებული შტატების სამხედრო დაზვერვა იყენებდა მეორე მსოფლიო ომის დროს ამ ენის წარმოდგენელი სირთულისა და მის მცოდნეთა რაოდენობის უკიდურესი სიმცირის გამო მთელ დედამიწაზე (ნახ 13).



ნახ. 13 მშიფრავი რადისტები ნავახოს (Navajo) ტომიდან,  
1944 წელი

2000 წლის 21 დეკემბერს ნავახოს ტომის მშიფრავებს მიენიჭა ამერიკის შეერთებული შტატების კონგრესის ოქ-

როს მედალი — ქვეყნის უმაღლესი სამოქალაქო ჯილდო. კონგრესის ოქროს მედლით დაჯილდოების საფუძველია გამორჩეული მიღწევები, რომლებსაც განსაკუთრებული მნიშვნელობა აქვს შეერთებული შტატების ისტორიისა და კულტურისათვის. 2014 წელს კი ნავახოს ტომის ინდიელებმა მიიღეს ხელისუფლებისაგან 500 მილიონ დოლარზე მეტი კომპენსაცია — უდიდესი, რომელიც კი ოდესმე გაცემულა *ავტოქთონების-თვის* (ძველისძველი მკვიდრი მოსახლეობისთვის, ანუ აბორიგენებისთვის) ამერიკის შეერთებულ შტატებში. თქვენ მოცემული გაქვთ სახელმძღვანელო, სადაც ახსნილია, როგორ უნდა მოხდეს იეროგლიფების კომბინირება; მაგრამ თავად ამ სიმბოლოების მნიშვნელობათა ახსნა არ ხდება იმ მიზეზით, რომ ეს საჭირო არ არის ფრაზების შესადგენად. ოთახის გარეთ იმყოფებიან ჩინური ენის მცოდნე დამკვირვებლები. მათ თვალწინ თქვენ სრულიად სწორად ადგენთ გარკვეული აზრის მქონე წინადადებას და მთლიან ტექსტებსაც, მაგრამ ამ დროს ეს აზრი თქვენთვის სრულიად გაუგებარია (იხილეთ ნახ. 14).



ნახ. 14 ჩინური ოთახი — აზრში წარმოდგენილი  
 ექპერიმენტი  
 (1980, John Rogers Searle)

შემდეგ დამკვირვებლები გადმოგცემენ წარწერებს შეკითხვებით ჩინურ ენაზე. თქვენ ადვილად შეგიძლიათ მათზე პასუხების გაცემა საჭირო სიმბოლოების შერჩევით (სახელმძღვანელოში მოცემული წესების საფუძველზე).

გარეგნულად პასუხები შეესაბამება შეკითხვებს და სავსებით გააზრებულია, თუმცა ისინი შეიძლება არ შეესაბამებოდეს რეალურ სინამდვილეს. მაგალითად, თქვენ შეგეკითხეს რომელი ფერი მოგწონთ ყველაზე მეტად

და თქვენ დაწერეთ «წითელი», მაგრამ თქვენ ხომ მწვანე მოგწონთ?!

სინამდვილეში თქვენ ხომ წარმოდგენაც არ გაქვთ იმის შესახებ, თუ რა კითხვა დაგისვეს ფერის შესახებ და გაეცით გარკვეული პასუხი — უბრალოდ შეათავსეთ ფუძე დაბოლოებასთან, აბსტრაქტული სუბიექტი აბსტრაქტულ ობიექტთან და ასე შემდეგ.

ძნელი არ არის იმის გაგება, რომ სისტემა, რომელიც შედგება ოთახისგან და მასში მყოფი ადამიანისგან — ეს კომპიუტერის მოდელია, სახელმძღვანელო — პროგრამა, ხოლო იეროგლიფების კრებულები — მონაცემთა ბაზა. ოთახის გარეთ მყოფ დამკვირვებლებს შეჰყავთ რაღაც მონაცემები, რომლებსაც «კომპიუტერი» ამუშავებს პროგრამის შესაბამისად, მაგრამ სრულიად არ ესმის, რა ინფორმაციაზე მუშაობს იგი კონკრეტულად და რისთვის არის ეს საჭირო. სხვანაირად რომ ვთქვათ, იგი დაკავებულია სინტაქსით — ობიექტთა (სიმბოლოთა) კომბინაციების შედგენით ნასწავლი წესების საფუძველზე, მაგრამ არ ესმის საკუთრივ სიმბოლოების და მათი კომბინაციების აზრი.

თქვენ რომ ჭამა მოგიხდეთ (ესე იგი «კომპიუტერს» მოუნდეს ჭამა), უდავოდ, ვერ შეძლებთ ამის თქმას დამკვირვებლებისთვის, რომც იცოდეთ ჩინური სინტაქსი. ე. ი. ინტელექტის მქონე მანქანასაც კი არ აქვს თავისი ნების გამოხატვის უნარი, რომც გააჩნდეს ეს ნება.

ეს ორი მდგენელი — მნიშვნელობათა არე (სემანტიკა) და საკუთარი ნების გამოხატვის უნარი — ადამიანის გააზრებული შემოქმედების მთავარი არსია. სინტაქსი ასევე აუცილებელია, მაგრამ იგი აზრთა გადაცემის მექანიზმის მხოლოდ დამხმარე როლს ასრულებს.

#### **IV. ადამიანი როგორც ბიოლოგიური რობოტი**

თავის ექსპერიმენტში ჯონ სირლი არა მხოლოდ აკრიტიკებს სრულფასოვანი ხელოვნური ინტელექტის შესაძლებლობას. «ჩინური ოთახი» ხსნის იმ ფორმალიზებული განათლების არსსაც, რომელიც თანდათანობით ინერგება სხვადასხვა ქვეყნის სკოლებსა და პროფესიულ სასწავლებლებში. ასეთი მიდგომა ტიპური ამოცანების სწრაფად დაძლევის საშუალებას ქმნის, მაგრამ არ ასწავლის იმის გაგებას, რაც და რატომ კეთდება. ასეთი

მიდგომით ნასწავლი ადამიანი ხდება თავისებური ბიოლოგიური რობოტი, რომელსაც შეუძლია დეტალების ვირტუოზულად შეერთება და მოქმედებათა მრავალსაფეხურიანი თანამიმდევრობის შესრულება, მაგრამ არ შეუძლია პროცესის მართვა, მისი შედარება უფრო მაღალი დონის ამოცანებთან, პრობლემათა გადაჭრისადმი შემოქმედებითი მიდგომა და, ზოგადად, იმის არსში წვდომა, რაც ხდება. *ამისათვის აუცილებელია ტრადიციული, ანალიტიკაზე აგებული, განათლება, დროში ხანგრძლივი და ინფორმაციულად გაჯერებული.*

## V. ექსპერიმენტის კრიტიკა

ჯონ სირლის «გონება, ტვინი და პროგრამები» («Minds, Brains, and Programs», 1980) წიგნის გამოსვლის შემდეგ, რომელიც «ჩინური ოთახის» აღწერას შეიცავდა, მის თეორიას გაუჩნდა არა მხოლოდ მომხრენი, არამედ მოწინააღმდეგეებიც.

ზოგიერთი მეცნიერი აღნიშნავდა, რომ სისტემას, რომელიც შეიცავს ოთახს, ადამიანს, იეროგლიფებსა და სა-



ხელმძღვანელოს, თავისთავად აქვს გონება, ხოლო ადამიანი მასში — (წმინდა მექანიკური სამუშაოს შემსრულებელი) — მხოლოდ ერთ-ერთი კომპონენტია. მონაცემთა ბაზაში (იეროგლიფების კრებულში) მოცემულია ამ ზეტვინის ცოდნათა მთელი მასა. თუმცა ამ მეცნიერთ თავად დავიწყებით, რომ სისტემამ ვერ უპასუხა შეკითხვას, რა ფერს ამლევს იგი უპირატესობას, მისი «პასუხის» გარეგნულად უზადო გამართულობის პირობებში. სისტემის მონაცემთა ბაზაში მოთავსებულ ცოდნას არ შეუძლია გავლენის მოხდენა ამ სისტემის ქცევაზე (მისი ქცევა მოწყვეტილია, აბსტრაქტირებულია ამ ცოდნისგან, სისტემა მხოლოდ მექანიკურად უკავშირებს ერთმანეთს ცოდნის ფრაგმენტებს და არ უფიქრდება შედეგებს).

მეცნიერთა მეორე ჯგუფს, პირიქით, ეჭვი ეპარება ადამიანის გონების — როგორც რაღაც უნიკალურის — თავად არსებობაში. მათი აზრით, ადამიანის ყველა გონებრივი პროცესი — მხოლოდ სინტაქსია, თუმცა წარმოდგენილად რთული: წესების განშტოებული სისტემის ფორმირება ასეული ათასობით წლის განმავლობაში

ხდებოდა, ადამიანის ცხოველური წინაპრების ჩათვლით, მაგრამ საფუძველში მაინც მონაცემთა ფართო ბაზა და ამ მონაცემთა რამდენიმე მისაწვდომი კომბინაცია დევს. ადამიანები და ცხოველები, ამრიგად, მხოლოდ ბიოლოგიური რობოტებია, რომლებიც მონაცემებით მანიპულირებენ სხვადასხვა დონეზე. მათ ტვინს შეუძლია როგორც სასარგებლო სამუშაოს უშეცდომოდ შესრულება, ასევე «ბაგის» (ინგლ. bug — ვირუსი; ტილი, მწერი) აკიდებაც, რომლის როლს ასრულებს რელიგია (რელიგიურ-რიტუალური ქცევა ცხოველებში და ფრინველებშიც აღინიშნება ისევე ხშირად, როგორც ადამიანებში, მაგალითად, ყვავების გონების სრულიად ახალი კვლევა გასაოცარი აღმოჩენით დასრულდა — ეს ფრთხილი ფრინველი არა მხოლოდ ჭკვიანია, არამედ ფლობს ცნობიერების გარკვეულ ფორმას, შეუძლია შეგნებულად აწმყოში გაიაზროს მის ირგვლივ არსებული გარემო. სხვა სიტყვებით რომ ვთქვათ, აქვს სუბიექტური განცდა); გარედან შესაძლებელია ამ ტვინის დაპროგრამებაც და მათ შორის მისი «*დავირუსებაც*» (იდეოლოგიით, პროპაგანდით; ცხოველებში კი — წვრთნით).

## VI. დაიპყრობს თუ არა სამყაროს მანქანები ?

სავარაუდოთ, «მანქანათა ჯგანყი» არასოდეს მოხდება. კომპიუტერული ინტელექტი მართლაც რომ ახდენდეს ადამიანის ინტელექტის მოდელირებას, მაშინაც კი ახლო (ადვილად წარმოსადგენ) მომავალში მოდელი აშკარად ვერ გაუსწრებს ბუნებრივ ტვინს და მასთან ჩამორჩენილი დარჩება.

ხოლო, თუ რომელიღაც ზემდლავრი რობოტი მოაწყობს «ჯანყს» და დაესხმება ადამიანებს, მაშინ იგი ადამიანის მიერ დაწერილი პროგრამის მხოლოდ მექანიკურად შემსრულებელი აღმოჩნდება და რაიმე სარგებელს თავისთვის ვერ მიიღებს.

## VII. ა. ნიუელისა და ჰ. საიმონის ჰიპოთეზა

ნიუელისა და საიმონის (ნახ 15) *ჰიპოთეზა ფიზიკური სიმბოლური სისტემის შესახებ* არის ალენ ნიუელისა (Allen Newell) და ჰერბერტ საიმონის (Herbert A. Simon) მიერ 1976 წელს ჩამოყალიბებული ვარაუდი იმის შესახებ, რომ «*ფიზიკურ სიმბოლურ სისტემას აქვს აუცილედ*

ბელი და საკმარისი საშუალებები ძირითადი ინტელექტუალური ოპერაციების განსახორციელებლად» (ინტელექტუალურ ოპერაციებად იგულისხმება ძლიერი ხელოვნური ინტელექტის მოქმედებები).



ნახ. 15 ალენ ნიუელი (1927-1992) და ჰერბერტ საიმონი (1919-2001)

სხვა სიტყვებით, იგულისხმება, რომ *სიმბოლური გამოთვლების გარეშე შეუძლებელია გააზრებულ მოქმედებათა შესრულება, ხოლო სიმბოლური გამოთვლების ჩატარების უნარი სავსებით საკმარისია გააზრებულ მოქმედებათა განხორციელების უნარის შესაძენად. ამრიგად, თუ ვივარაუდებთ, რომ ადამიანი ან ცხოველი ან*

მანქანა მოქმედებს გააზრებულად, ისინი რაღაცნაირად ასრულებს სიმბოლურ გამოთვლებს. და პირიქით, რაკი კომპიუტერს შეუძლია მსგავსი გამოთვლების განხორციელება, ამიტომ მის საფუძველზე შესაძლებელია ხელოვნური ინტელექტის შექმნა. ჰიპოთეზას საფუძველად დაედო ნიუელისა და საიმონის მიერ შექმნილი პროგრამის — ამოცანათა უნივერსალური ამომხსნელის — გამოყენება ადამიანის მსჯელობათა მოდელირებისათვის. ჰიპოთეზის დაცვა კრიტიკისგან შეუძლებელია, მაგრამ ხელოვნური ინტელექტის კვლევათა უდიდესი ნაწილი სწორედ სიმბოლურ სისტემათა შექმნის გზით წავიდა. დამოუკიდებლად იმისა, სამართლიანია თუ არა ეს ჰიპოთეზა, სიმბოლური გამოთვლები — დაპროგრამების რეალობაა და მათი პრაქტიკული სარგებელი საზოგადოდ აღიარებულია.

## VIII. რჩეული მაგალითები

ინფორმაციის მასობრივი საშუალებები და ექსპერტები ყოველდღიურად გვიკვიჩინებენ, რომ ხელოვნური ინტელექტი მომდგარია კარს და სულ მალე ჩაგვანაცვლებს.

მაგრამ ის ხელოვნური ინტელექტი, რომლის შესახებ გვესაუბრებიან, სინამდვილეში არ არსებობს.

აქ ადგილი აქვს გაუგებრობას, რომელიც სათავეს იღებს ჯონ მაკარტისა და მისი სამი მეგობრის მიერ ამ ტერმინის შექმნის მომენტიდან 1956 წლის ცნობილ დარტმუთის კონფერენციაზე. ამ კონფერენციის შემდეგ ფრენკ როზენბლატმა 1957 წელს შექმნა სწავლების ალგორითმი, რომელიც პრეტენზიას აცხადებდა ნეირონის მუშაობის მაქსიმალურად შესაძლო მოდელირებაზე.

თუ შეიძლება ნეირონის მოდელირება, მაშინ რატომ არ განვახორციელოთ ნეირონული ქსელის მოდელირებაც? ამრიგად, ჩვენ შევძლებთ ხელოვნური ინტელექტის შექმნასაც.

ეს სულელური მსჯელობაა, რომელიც საფუძვლად უდევს იმას, რასაც დღეს ხელოვნური ინტელექტი ეწოდება.

თქვენ რომ დაგისვან კითხვა, რამდენი იქნება 1789+522, უკეთეს შემთხვევაში დაგჭირდებათ რამდენიმე წამი მაინც, რომ შედეგი გვითხრათ. და ამასთან ერთად, არის შანსი, რომ თქვენ ცდებოდეთ.

მეორე მხრივ, მანქანას შეუძლია ამ გაანგარიშების შესრულება მყისიერად და შეცდომის რისკის გარეშე. პრობლემა მხოლოდ ისაა, რომ ეს ინტელექტი არ არის. ამას მოწმობს იმ კვლევათა უშედეგო ფინალი, რომლებიც ჩატარდა დარტმუთის კონფერენციის შემდეგ. რა მოხდა 1960-იან წლებში? ასეთი მათემატიკური მიდგომის საფუძველზე განხორციელდა ბუნებრივი ენის მოდელირების მცდელობა. სხვა სიტყვებით რომ ვთქვათ, ყველაზე რთული საქმე ამ ქვეყნად. შედეგი ნულოვანი აღმოჩნდა. ბუნებრივი ენის პრობლემის გადაწყვეტის ეს მცდელობა კომპიუტერის საშუალებით, წარუმატებელი აღმოჩნდა.

ამის შემდეგ დადგა ხელოვნური ინტელექტის ეგრეთ წოდებული «პირველი ზამთარი»: შეწყდა დაფინანსება და, ამრიგად, შეწყდა კვლევებიც. მოცემული იყო ზღაპრული დაპირებები, ისინი შეუსრულებელი დარჩა, კვლევებიც შეჩერდა. ეს პირქუში, ბნელი პერიოდი — ერთ-ერთი მიზეზია, რომელიც უნდა აიძულებდეს დღევანდელ მკვლევართ თავიანთ მოვალეობად ჩათვა-

ლონ კალმის აღება: უნდა გამოაშკარავდეს იმათი სიცრუე, ტყუილი, რომელსაც ისინი ავრცელებენ სისულელეთა სახით ხელოვნურ ინტელექტზე. და ეს აუცილებელია თუნდაც იმიტომ, რომ თავიდან ავიცილოთ კვლევათა ახალი შეწყვეტა. ცხადია, სასარგებლო საქმეები შესაძლებელია, მაგრამ დაუშვებელია, რომ მკერდის კიბოს ან გზებზე სიკვდილიანობის საწინააღმდეგო საშუალებათა შექმნა დაზარალებს ხელოვნური ინტელექტის ახალ წინასწარმეტყველთა შებოლვით და სარკით მანიპულირების გამო.

ხელოვნური ინტელექტის ამ პირველი ზამთრის მიუხედავად, 1970-იან წლებში ადამიანები განაგრძობდნენ ცეცხლის შენარჩუნებას საექსპერტო სისტემათა იდეის ჩამოყალიბებით და განვითარებით. ეს სისტემები ასევე ეფუძნება ლოგიკასა და მათემატიკას, იყენებს წესებს. ასეთი საექსპერტო სისტემების ძალიან კარგი მაგალითია ჭადრაკი. 1997 წელს Deep Blue სისტემამ მოუგო გარი კასპაროვს ჭადრაკში. კომპიუტერი გახდა უფრო ინტელექტუალური, ვიდრე ადამიანი.



და განა ასეა? სინამდვილეში საქმე მაინც არ არის ინტექტში. ბოლოს და ბოლოს ეს მხოლოდ მანქანის სწავლებაა: იგი ეხება ჭადრაკის წესებს, რომლებიც ძალიან მარტივია მოდელირებისათვის და მანქანის მეხსიერებაში რაც შეიძლება მეტი რაოდენობის პარტიათა შეყვანას (კანონიერი პოზიციების საერთო რიცხვი შეადგენს  $10^{43}$ -დან  $10^{50}$ -მდე) იმისათვის, რომ ასეთსა და ასეთ კონფიგურაციაში მანქანამ შეძლოს საუკეთესო სვლის პოვნა მონაცემთა თავის ბაზაში. ეს ხომ უხეში, ბრუტალური ძალის პროცესის გამოყენებაა? მანქანამ გაიმარჯვა, იმიტომ რომ თავისი მეხსიერების გამო მან შეძლო წინასწარ უფრო მეტი სვლის პროგნოზირება, ვიდრე კასპაროვმა. 1990-იანი წლების სხვა ტენდენცია დაკავშირებული იყო ნეირონულ ქსელებთან, რამაც მიგვიყვანა დღევანდელ *მანქანურ სწავლებამდე*. ამ კვლევითი პროგრამის ნიშანსვეტი ინტერნეტი გახდა. რატომ მოხდა ასე? იმიტომ, რომ ინტერნეტი — მონაცემთა ყველაზე დიდი ბაზაა, რომელიც კი შეიძლება წარმოიდგინოს დღეს ადამიანმა. ხოლო *მანქანური სწავლება* მოითხოვს მონაცემ-

თა დიდ რაოდენობას. თუ *ღრმა სწავლების* განვითარება 2000-იან წლებში დაიწყო, ეს მხოლოდ ინტერნეტის წყალობით მოხდა.

ავიღოთ ცნობილი მაგალითი, რომელიც კატებს ეხება. ყველგან ამ პლანეტაზე უამრავი ადამიანი ათავსებს თავისი კატის ფოტოსურათს ინტერნეტში. შედეგად, რამდენიმე წლის განმავლობაში შეიქმნა მონაცემთა განსაცვიფრებელი ბაზა კატების რამდენიმე მილიონი ფოტოსურათით, რაც მკვლევარებს მძლავრ ინსტრუმენტს აძლევს მათი მანქანების მოსამზადებლად ნიმუშების — ამ შემთხვევაში, კატების — ამოცნობისათვის. ამჟამად არის ისეთი ალგორითმების შექმნის შესაძლებლობა, რომლებიც გამოიყენებს კატათა 100 000 გამოსახულებას და შემდეგ ამოიცნობს კატას 98% სიზუსტით. ცხადია, ეს გარემოება შეიძლება შთამბეჭდავ ფაქტად მოგვეჩვენოს. სინამდვილეში კი ეს ძალიან ცუდი მახასიათებელია. კატის რამდენი გამოსახულება იქნება საჭირო ადამიანისთვის, რომ მან იგივე გააკეთოს? ორი ... სამართლიანობისთვის უნდა ვაღიაროთ, რომ ადამიანების

შემთხვევაში ჩვენ ვლაპარაკობთ კატის ორ ეგზემპლარზე (ცალზე), და არა უბრალოდ სახეზე.

მაგრამ ეს ნიშნავს, რომ ჩვენი ინტელექტი მულტიმოდალურია:

იგი იყენებს არა მხოლოდ იმას, რასაც თვალი ხედავს, არამედ გრძნობათა სხვა ორგანოსაც და მრავალ ისეთ რამეს, რომლის მოდელირება გამოსახულების მოდელირებაზე გაცილებით უფრო ძნელია.

ეს იმას ნიშნავს, რომ ეგრეთ წოდებულ ხელოვნურ ინტელექტს და ადამიანის ინტელექტს ერთმანეთთან საერთო ცოტა რამ აქვს.

მეორე მაგალითია თამაში «გო».

20 წლის შემდეგ Deep Blue ჭადრაკის მოთამაშე კომპიუტერის გამარჯვებიდან გარი კასპაროვზე, მსოფლიოს ჩემპიონი გოს თამაშში (ნახ. 16), სამხრეთ კორეელი ლი (გვარი) სედოლი (სახელი) — Lee Sedol — 2016 წელს, დამაჯერებელი (1 : 4) ანგარიშით დამარცხდა ხუთპარტიან ჭიდილში AlphaGo პროგრამასთან, რომელიც Google DeepMind-მა შექმნა.



ნახ. 16 Lee Sedol (1983-) მე-9 დანის კორეელი გო-  
პროფესიონალი (სამხრეთ კორეა)

«გო» — ჭადრაკზე ბევრად უფრო რთული თამაშია იმ თვალსაზრისით, რომ აქ ნებადართულია დაახლოებით  $10^{170}$  ლეგალური პოზიცია. ეს წარმოუდგენლად ალემატება შესაბამის რიცხვს ჭადრაკში. შეიძლება იმის თქმაც კი, რომ საქმე გვაქვს უსასრულობასთან. შედარებისათვის საკმარისი იქნებოდა თუნდაც იმის აღნიშვნა, რომ ამ სამყაროში მოლეკულათა რაოდენობა მხოლოდ  $10^{180}$  რიცხვს შეადგენს. სხვა სიტყვებით რომ ვთქვათ, არანაირი მეხსიერების შესაძლებლობაშია ამოიცნოს ყველა

სავარაუდო სვლა ან თუნდაც მათი მიახლოებითი რაოდენობა ამ თამაშში. ამიტომ მანქანის პროგრამისტებმა სხვა სტრატეგია აირჩიეს: ნაცვლად იმისა, რომ მიეცათ პროგრამისტვის სვლათა ბიბლიოთეკა, ასწავლეს მას შემოეფარგლა მოქმედება 30 000 თამაშის (პარტიის) ნაკრებზე. ეს ნაკრები ძალიან მცირეა შესაძლო თამაშთა რაოდენობის მიმართ, მაგრამ AlphaGo პროგრამამ, ამ რჩეული თამაშებისა და მათი ანალიზის წყალობით, შეიძინა უნარი, რომელმაც საშუალება მისცა მას არა მხოლოდ მოეგო მსოფლიოს ჩემპიონისთვის, არამედ შემოეთავაზებინა განსაცვიფრებელი სვლები. გოს რამდენიმე უდიდესი თანამედროვე მოთამაშე იმასაც კი ამბობს, რომ ამ სვლებმა შეცვალა კიდევ თამაშის მათი საკუთარი კონცეფცია.

წარმადობა, მართლაც, შთამბეჭდავია, მაგრამ ამ შედეგის მისაღებად Deep Mind კომპანიას მოუხდა ისეთი «წვრილმანის» მობილიზება, როგორიცაა 1 500 CPU (Central Processing Unit, ცენტრალური პროცესორი), 300 GPU (Graphics Processing Unit, გრაფიკული პროცესორი)

და 30 TPU (Tensor Processing Unit, ტენზორული პროცესორი), ორი ათასი პროცესორი და მათ შორის რამდენიმე ასეული მრავალბირთვიანი «ბეჰემოთი». ეს მონაცემთა დამუშავების ნამდვილი ცენტრია, რომელიც 440 kW·h (კვტ·სთ) ელექტროენერგიას მოიხმარს, რაც მეტია, ვიდრე ადამიანს ესაჭიროება, სახელდობრ, მხოლოდ 20 W·h (ვტ·სთ)-დან 25 W·h (ვტ·სთ)-მდე. სხვა სიტყვებით, მანქანის შესაქმნელად, რომელიც დაამარცხებს ადამიანს გოს პარტიში, — და მხოლოდ იმიტომ, რომ ეს კომპიუტერი ზესპეციალიზებულია — საჭიროა ენერგიის კოლოსალური რაოდენობა, მაშინ როცა ადამიანი მხოლოდ რამდენიმე ვატ·საათს მოიხმარს.

*მანქანური სწავლების დამოკიდებულება მონაცემებზე ასევე არის შეცდომათა ძირითადი წყარო. ალგორითმმა შეიძლება შეცდომა დაუშვას არა მხოლოდ ხარვეზების გამო დაპროგრამებაში, არამედ შეცდომათა მიზეზითაც მონაცემებში. 2016 წელს Microsoft კომპანიამ აამოქმედა ჩატ-ბოტი Tay (ფრაზის აკრონიმი: «Thinking About You» — «შენზე ფიქრში»). იგი უნდა შესულიყო ურთიერთობებში მომხმარებლებთან და ხელი შეეწყო Microsoft-ის*

დამკვიდრებას Twitter-ში, რომელიც 2022 წლის მიწურულს ამერიკელმა მეწარმემ, ინჟინერმა და მილიარდერმა იტონ რივ მასკმა (Elon Reeve Musk, 1971) შეიძინა 44 მილიარდ ამერიკულ დოლარად. მაგრამ მოულოდნელად პრობლემა გაჩნდა: სულ რაღაც თექვსმეტი საათის განმავლობაში Tay გახდა ყველაზე ცუდი *ტროლი-რასისტი*, რომელიც კი უნახავთ ოდესმე ისტორიაში. კომპანიის «*ტვინის ცენტრმა*» არა ერთი თვე მოანდომა იმ რობოტის შექმნას, რომელიც ახლა სასწრაფოდ უნდა გაეთიშათ ქსელიდან მუშაობის დაწყებიდან სულ რამდენიმე საათის შემდეგ. რა მიზეზი აქვს ასეთ მარცხს? კორექტირების შეცდომა : *ჩატ-ბოტები* განკუთვნილია ინტეგრაციისათვის თავისი მიზნის განსახორციელებლად. მაგრამ აღმოჩნდა, რომ Twitter-ში ადამიანები შეურაცყოფას აყენებენ ერთმანეთს, ვრცელდება ქსენოფობიური კომენტარები, მრავალი მომხმარებელი დიდი სიამოვნებით უთმობს დროს Tay რობოტის «აღზრდას». ეტყობა, რობოტის შეწყობადობის, შეგუებადობის ფაქტორი მოითხოვდა შემცირებას. მაგრამ შეცდომათა კი-

დევ ერთ წყაროს წარმოადგენს Tay რობოტის სწავლებისთვის შერჩეული მონაცემების ბაზა. ვინაიდან იმისათვის, რომ რობოტს შეასწავლო ურთიერთობა საუბრისა და აზრის გაცვლის ფორმით, მას უნდა ჩაუტარდეს ტრენინგი მრავალ ათას რეალურ საუბარზე ადამიანთა შორის. მოხდა ისე, რომ არსებობს მონაცემთა *Switchboard* ბაზა. რომელიც შეიცავს საუბრების გაშიფვრას კლიენტებსა და კოლ-ცენტრებს (ინგლ. call center — ზარების დამუშავების ცენტრი) შორის ამერიკის შეერთებულ შტატებში დაწყებული 1950 წლებიდან. ეს ფასდაუდებელი ინსტრუმენტია ლინგვისტებისთვის. იმის გამო, რომ მონაცემთა ბაზა ძალიან დიდია მისი მთლიანი მოცულობით გამოყენებისთვის, შეიძლება წარმოვიდგინოთ, რომ Tay რობოტის პროგრამისტებმა აირჩიეს მონაცემთა ამ ბაზის ქვესიმრავლე. ვივარაუდოთ, რომ ამ პროგრამისტებმა აიღეს ლექსიკოგრაფიულად პირველი შტატის პირველივე პერიოდის საუბრები. სხვა სიტყვებით, *ალაბამა 1950-იან წლებში ...* საშინელებაა! მონაცემთა ცუდი არჩევა შეიძლება კატასტროფული არაობიექტურობის მიზეზი გახდეს.



ხელოვნური ინტელექტი — ეს არ არის ინტელექტი. ეს ასევე არ არის «შავი ყუთი», რომელიც არ გვექვემდებარება. ხელოვნური ინტელექტი არის ის, რასაც მთლიანად ვაკონტროლებთ. თუ ჩნდება შეცდომები, ეს ხდება შეცდომათა გამო ალგორითმებში ან მონაცემებში. და ეს შეცდომები ყოველთვის შეიძლება იყოს ახსნილი. თეორიულად ხელოვნურ ინტელექტში ყველაფერი ახსნადია. პრაქტიკულად კი ყველაფერი შეიძლება უფრო რთული აღმოჩნდეს. ჩვენ ვლაპარაკობთ მანქანებზე, რომლებიც აწარმოებს რამდენიმე მილიონ გამოთვლას წამში. თუ მოგვიწევს უკან დაბრუნება და ყველა იმ განგარიშების გადახედვა, რომელიც მანქანამ გააკეთა, ესეც დაიკავებს რამდენიმე მილიონ წამს, რომელიც არ გვაქვს. ამიტომ შეიძლება ვილაპარაკოთ ამოუხსნადობის შესახებ, მაგრამ მხოლოდ პრაქტიკული მოსაზრებებიდან გამომდინარე. მათემატიკის თვალსაზრისით კი, ყველაფერი, რასაც აკეთებს ეგრეთ წოდებული ხელოვნური ინტელექტი, სავსებით ახსნადია.

2017 წელს დაიწყო ლაპარაკი იმის შესახებ, რომ ავტონომიური ავტომობილი სულ მალე გამოჩნდებოდა. ორი

წლის შემდეგ ვრცელდებოდა აზრი, რომ ამისათვის 5 ან 10 წელი იქნება საჭირო და ეს აუცილებლად მოხდება. კიდევ ერთი წლის შემდეგ ავტოპილოტირებული ავტომობილის გამოჩენის თარიღი ოცი წლით გადაიდო. სინამდვილეში მეხუთე დონის ავტონომიური ავტომობილი, რომელიც თავად შეძლებს ნებისმიერ სიტუაციაში ყველაფრის გაკეთებას, არც ოცი, არც ასი, და არც ათასი წლის შემდეგ გაჩნდება : საერთოდ არასოდეს.

დავუშვათ, რომ ვართ პარიზში ავტომობილით შარლ დე გოლის მოედანზე (Place Charles de Gaulle), რომელიც ისტორიულად ცნობილია როგორც ვარსკვლავის მოედანი (Place de l'Étoile), და საღამოს ექვსი საათია. რაგინდ სიღრმისეულად უნდა იყოს თქვენი ავტომობილის მიერ შესწავლილი საგზაო მოძრაობის წესები, იგი თავს ვერ გაართმევს სიტუაციას, ვინაიდან ვარსკვლავის მოედანი ამ დროს — ეს არა მხოლოდ საგზაო მოძრაობის წესებია, არამედ სოციოლოგიაც და მოლაპარაკებებიც.

ცხადია, რომ ერთიცა და მეორეც შეიძლება იყოს მოდელირებული, რაც, ვალიაროთ (და ჩავხედოთ სიმართლეს

თვალეებში), დიდ დროს წაგვართმევს. ახლა ჩავსხდეთ ჩვენს ავტომობილში, რომლის აწყობას მრავალი წელი მოვანდომეთ, და გავემგზავროთ აზიის ყველაზე დინამიურად განვითარებად უმსხვილეს ქალაქ *ბენგალურუში* (Bengaluru), რომელიც «ინდოეთის სილიკონის ველად» მოიხსენიება ...

მრავალწლიანი მუშაობა იქნება საჭირო ისეთი ავტომობილის გასაკეთებლად, რომელიც შეძლებს გადარჩენას არა მხოლოდ პარიზის მოედანზე *Étoile*, არამედ ამ ბენგალურუშიც, რომელიც, ხატოვნად რომ ვთქვათ, *სულ სხვა ფერის კარტის წყვილია*. ადამიანის ინტელექტს შეუძლია ადაპტირება ახალ ვითარებასთან და გამომგონებლობის გამოჩენა ყოველგვარი სპეციალური მომზადების გარეშე ამა თუ იმ კონკრეტული სიტუაციისთვის. მანქანას კი, პირიქით, მსგავს შემთხვევაში მოქმედების გამოცდილება ადრე უნდა მიეღო.

რატომ ვახსენეთ კარტი და მისი ფერი? *ბელოტის* სახელწოდებით ცნობილ კარტის (ბანქოს) თამაშში, რომელიც ძალიან გავრცელებულია დასავლეთ ევროპაში (განსაკუთრებით, საფრანგეთში), გვხვდება ტერმინი

*რებელოტი (Rebelote) — მეფის ჩამოსვლა მისივე ფერის ქალის ჩამოსვლის შემდეგ. ფრანგულ ხალხურ მეტყველებაში ამ სიტყვას მეორე — იდიომის — გაგებითაც ხმარობენ: ყველაფერი იგივე! სწორედ ამ გაგებით ვიხმარეთ ეს სიტყვა აქ გადმოცემული სიტუაციის დასახასიათებლად: პარიზი და ბენგალურუ არ არის ბელოტი და რებელოტი!*

საილუსტრაციოდ მშვენიერი მაგალითი იქნება Google-ის ქვედანაყოფი Waymo LLC (Limited Liability Company — შეზღუდული პასუხისმგებლობის საზოგადოება), რომელიც დაკავებულია ავტომობილების ავტონომიური ტარების პრობლემით. 2020 წელს კომპანიამ გააფართოვა მოსახლეობის ხსენებული მომსახურება და ახლა ეს ერთადერთი კომერციული სერვისია უპილოტო ავტომობილებისთვის. იგი მუშაობს სარეზერვო მძღოლების გარეშე მანქანაში. საინტერესოა, რომ უკანასკნელი ათი წლის განმავლობაში Waymo-ს გავლილი აქვს ავტონომიურ ავტომობილებზე თვრამეტი მილიონი კილომეტრი. თუმცა, ამის მიუხედავად, მისმა გენერალურმა დირექტორმა ჯონ კრაფციკმა (John Krafcik,

1961) დაუმალავად განაცხადა 2019 წლის მარტში, რომ მე-5 დონის ავტონომიური ავტომობილი არასოდეს იარსებებს. საინტერესოა, რომ მაშინვე ინტერნეტში ღია წვდომის უზრუნველყოფით მან განათავსა თავისი კომპანიის წვრთნის (ტრენინგების) ვიდეოჩანაწერები. ცხადია, რომ, მასალის სპეციფიკიდან გამომდინარე, ეს პროგრამა სახალისო ვერ იქნებოდა. იგი მოსაწყენი დარჩებოდა, რომ მასში განსაცვიფრებელი წვრილმანი არ აღმოჩენილიყო.

ამ ვიდეოში საქარე მინის (ინგლ. Windshield, ფრანგ. le pare-brise, pare — ზმნიდან parer: მოგერიება, აცდენა..., briser — მსხვრევა) უკან განთავსებული კამერა აჩვენებს იმას, რასაც ხედავს მძღოლი. ავტომობილი გადაადგილდება და შემდეგ უცებ ჩერდება. მალე იგი კვლავ იწყებს გადაადგილებას, მაგრამ ისევ ჩერდება და ასე მრავალჯერ, ვიდრე ოპერატორი მიხვდებოდა თავად აელო ხელთ საჭე და ემართა მანქანა. რამ გამოიწვია ავტომობილის ასეთი უცნაური ქცევა? კადრების ყურადღებით შესწავლისას შეინიშნება ბორდიურზე მიმავალი ორი

ადამიანი. ერთ-ერთს მხარზე გადაკიდებული აქვს ჩან-  
თა, საიდანაც «სდექ» საგზაო ნიშნის თავია ამოყოფილი.  
როცა საჭესთან ადამიანია, მას ამ სურათზე რეაქცია არ  
ექნება და არც რაიმეს მოიმოქმედებს. მაგრამ მანქანას  
მსგავს სცენართან საქმე არასოდეს ჰქონია. ამიტომ იგი  
იმ დასკვნამდე მივიდა, რომ იყო «სდექ» საგზაო ნიშანი  
და გაჩერდა კიდევ. ცხადია, რომ მსგავსი სცენარები  
შეიძლება გავითვალისწინოთ, დავგვეგმოთ და მანქანას  
სათანადო სწავლება ჩავუტაროთ. მაგრამ, გარდა იმისა,  
რომ ეს პრობლემა ახალ თავსატეხს გაგვიჩენდა, უფრო  
მნიშვნელოვანი პრობლემა იმაში მდგომარეობს, რომ  
შეუძლებელია წინასწარ ყველა ანომალიის პროგნოზი-  
რება, რადგან მათი რაოდენობა პრაქტიკულად უსასრუ-  
ლო სიმრავლეს შეადგენს. ჩვენ, ადამიანებსაც, არ შეგვი-  
ძლია ყველა შემთხვევის წინასწარმეტყველება, მაგრამ  
უპრეცედენტო მოვლენასთან პირისპირ შეჯახებისას  
ადამიანის ინტელექტს *გამომგონებლობის უნარი* აქვს.  
ეს ხომ არ არის სტატისტიკაზე და ლოგიკაზე, წესებსა  
და მონაცემებზე დაფუძნებული ეგრეთ წოდებული «*ხე-  
ლოვნური ინტელექტი*»?!

კიდევ ერთი საინტერესო მაგალითი, რომლის შესახებ ინფორმაცია მოიპოვება, ეს არის Siri (Speech Interpretation and Recognition Interface — მეტყველების ინტერპრეტაციისა და ამოცნობის ინტერფეისი) — ღრუბლოვანი პერსონალური ასისტენტი და კითხვა-პასუხის სისტემა, პროგრამული კლიენტი, რომელიც Apple კომპანიის მრავალ პროდუქტში შედის. ხსენებული აპლიკაცია იყენებს ბუნებრივი მეტყველების დამუშავებას იმისათვის, რომ პასუხი გაეცეს კითხვებს და ჩამოყალიბდეს რეკომენდაციები. Siri სისტემას შეუძლია შეეწყოს, და შეეგუოს ყველა მომხმარებელს ინდივიდუალურად, მისწავლოს მისი უპირატესობები ხანგრძლივი დროის განმავლობაში ...

როცა 1997 წლიდან დაიწყო მუშობა სისტემაზე Siri, რაღაც მომენტში კონსტრუქტორთა შემოქმედებითი ჯგუფის წინაშე აღიმართა გადაულახავი ჯგერბი: ხმის ამოცნობი ამუშავდა, მაგრამ მხოლოდ 80%-ით. შეიძლება მოგვეჩვენოს კიდევ, რომ ეს საკმარისია, მაგრამ სინამდვილეში მაჩვენებელი დაბალია. მართლაც, ეს ხომ იმას ნიშნავს, რომ მარცხი, წარუმატებლობა 20%-ს შეადგენს

— სამოც სიტყვას სამას სიტყვიან გვერდზე! როცა ასეთ ფაქტს წააწყდნენ, გადაწყდა «*ლამის კლუბის პარადიგმის*» გამოყენება. საქმე ისაა, რომ ლამის კლუბში მუსიკა ფარავს თქვენი მეზობლის სიტყვებს. იგი, ალბათ, ნასვამია (გამორიცხული არ არის, თქვენც, სხვათა შორის). ამის შედეგად თქვენ თითქმის ვერაფერს იგებთ მის ნაღაპარაკებიდან, მაგრამ ამას არ ამჟღავნებთ და პასუხობთ მას მეტ-ნაკლებად ბუნდოვნად, გაურკვეველად. ეს სწორედ ისაა, რისი გაკეთებაც მოხერხდა დამპროექტებლების მხრიდან: ისინი აზროვნებდნენ «*ხელოვნური სისულელის*» და არა *ინტელექტის* ტერმინებში. რატომ? დამპროექტებლებმა დაიწყეს იმით, რომ Siri არც თუ ისე ჭკვიანია, ბევრი რამ არც ესმის, მაგრამ სავსებით მისაღებია საუბრის გასაგრძელებლად ხუმრობებით, სასაცილო ამბებით, ანეკდოტებით, შეკითხვებით... სხვა სიტყვებით რომ ვთქვათ, დამპროექტებლებმა დაიწყეს საქმე ხელოვნური სისულელით, რომ პროგრამა ეფექტური გამხდარიყო.



ამიტომ, ხელოვნური ინტელექტის ნაცვლად, უკეთესია ვილაპარაკოთ *შემავსებელ, დამატებით ანუ კომპლემენტარულ ინტელექტზე* (CI — Complementary Intelligence). ეს ჯერ კიდევ AI (Artificial Intelligence) ხელოვნური ინტელექტია, მაგრამ სხვანაირი : მეცნიერება არ აცხადებს პრეტენზიას ისეთი ავტონომიური სუბსტანციის შექმნაზე, რომელიც შეძლებს არსებობას ადამიანის გარეშე; მეცნიერება ქმნის *ინსტრუმენტებს* ადამიანის ინტელექტის ასამაღლებლად, რათა დაეხმაროს ადამიანს აკეთოს ის, რაც მას უკეთ გამოუდის.

როგორც ნებისმიერი ინსტრუმენტის შემთხვევაში, ჩვენ შეგვიძლია ბოროტადაც გამოვიყენოთ ისინი. განა ჩაქუჩი არ შეიძლება გამოვიყენოთ არა მხოლოდ ლურსმნის ჩასაჭედებლად, არამედ მეზობლის თავის გასატეხად? კანონები არეგულირებს გამოყენებათა ამ ფორმებს და მათ შორის ზოგიერთს საერთოდ კრძალავს. ცხადია, რომ ხელოვნური ინტელექტის რეგულირებაც უნდა ხდებოდეს ნებისმიერი სხვა ინსტრუმენტის მსგავსად, რომ თავიდან იყოს აცილებული მისი არასწორი მოხმარება. ამას დიდი მნიშვნელობა აქვს: გადაწყვეტილებას

ხომ მხოლოდ ადამიანები იღებენ?! მხოლოდ ადამიანს აკისრია პასუხისგებლობა ხელოვნური ინტელექტის არასწორ გამოყენებასთან დაკავშირებით და მხოლოდ მას შეუძლია გადაწყვეტილების მიღება ასეთი გამოყენების რეგულირების თაობაზე. სხვა სიტყვებით რომ ვთქვათ, ინტელექტუალურ ინსტრუმენტს არ აქვს შესაძლებლობა დამოუკიდებლად აკეთოს რაიმე. თუ ხელოვნური ინტელექტი აკეთებს რაიმე დასაგმობს, ეს მხოლოდ იმიტომ, რომ ამის დავალება ადამიანისგან მიიღო. სიტუაციას აკონტროლებს ადამიანი და მის გარეშე არაფერი ხდება.

ჩვენი გონება უსასრულოა: შეგვიძლია ორიენტირება იმაში, რაც ჯერ არ არსებობს, მაგრამ უახლოეს წლებში გაჩნდება; შეგვიძლია ორიენტირება იმაშიც, რის გამოგონებასაც მოვახერხებთ.

სწორია, რომ ჭადრაკში მანქანა ჩვენზე ძლიერია, მაგრამ ნაკლებად სწორი იქნება ასეთი შეფასება გოსთან მიმართებაში, ხელოვნური ინტელექტის პრობლემების გადა-

წყვეტიტ დაკავებული DeepMind ბრიტანული კომპანიის ტრიუმფის მიუხედავად, თუ ჩვენ გავითვალისწინებთ ენერჯის დანახარჯს.

მოიხმარო 440 კვტ.სთ თამაშის ერთ პარტიაში ადამიანის დასამარცხებლად, რომელსაც მხოლოდ 25 ვტ.სთ ელექტროენერჯია ესაჭიროება, ესოდენ საინტერესოა? ხოლო ისეთი პრობლემისათვის, როგორც ავტომობილის ტარებაა, არც ერთი პროგრამა არასდროს იქნება მთლიანად ავტონომიური.

ცხადია, ამჟამინდელი ტექნოლოგიური მიღწევები ძალზე საინტერესოა. მაგრამ რამდენად სწორია ლაპარაკი «ხელოვნურ ინტელექტზე» ამ მიღწევებისათვის კვალიფიკაციის მიცემისას? საზიანოა, უარყოფითი შედეგის მომტანია ლაპარაკი მთლიანად ავტომატურ მანქანებზე, რომლებმაც ოდესღაც (ეს ოცნებაა, თუ საშინელი სიზმარი?!) უნდა ჩაანაცვლოს — *სოციალური ცხოველისა და ერთადერთი ორფეხა პრიმატის* — ადამის ჯილაგი...

მეხუთე დონის თვითმართვადი ავტომობილი, როგორც ვთქვით, არასოდეს ივლის, მაგრამ მეოთხე დონის ავტომობილი ერთხელ გაჩნდება და ეს წინგადადგმული

ნაბიჯი იქნება: იგი არ სვამს, სიგარეტსაც არ ეწევა, ყურადღებასაც არ ადუნებს, ფოკუსირდება მხოლოდ იმაზე, რაც უნდა აკეთოს. ამას კი შეუძლია მნიშვნელოვნად შეამციროს სიკვდილიანობა გზებზე საავტომობილო კატასტროფებში. ამ მიმართულებით კვლევათა გაგრძელება აუცილებელია. მაგრამ ხელოვნური ინტელექტის *მირაჟმა* არ უნდა აგვაცდინოს გზას, როგორც ეს ხდება ზოგჯერ. მაგალითად, ჟურნალ Nature-ში 2019 წლის 7 ნოემბერს გამოქვეყნებულ პრეს-რელიზში საუბარია იმაზე, რომ *«ნეირონული ქსელები შეიძლება იყოს გამოყენებული ფიზიკური კონცეფციების აღმოსაჩენად ყოველგვარი წინასწარი ცოდნის გარეშე»*. ეს ინფორმაცია ღიმილს იწვევს, ვინაიდან გელიოცენტრიზმის პრინციპი ცნობილია ჯერ კიდევ სამყაროს გელიოცენტრული სისტემის — (პირველი სამეცნიერო რევოლუციის გარიჟრაჟის!) — ავტორის, გერმანელი ასტრონომის, მათემატიკოსის, მექანიკოსისა და ეკონომისტის ნიკლას კოპერნიკის (გერმ. Niklas Koppernigk, 1473-1543) დროიდან ექსპერიმენტული მონაცემების წყალობით და მარტივი ობიექტური დაკვირვებების საფუძველზე: მზე უნდა

იყოს განთავსებული მზის სისტემის ცენტრში ვარსკვლავისა და პლანეტა მარსის მოძრაობის შესაბამისად, როგორც ეს დედამიწიდან ჩანს.

ამ პიროვნების გერმანელად მოხსენებასთან დაკავშირებით უნდა ითქვას, რომ კოპერნიკის ეთნიკური კუთვნილების საკითხი საკამათოა. დედა, ბარბარა ვატცენროდე (Barbara Watzenrode, circa 1440 – circa 1495), გერმანელია, ხოლო მამის ეროვნება დღემდე დაუდგენელია და მხოლოდ ის არის ცნობილი, რომ იგი დაბადებული იყო კრაკოვში. მამის ძალიან ნაადრევი სიკვდილის შემდეგ, კოპერნიკი იზრდებოდა დედისა და ბიძის — ეპისკოპოსის, დიპლომატის, კანონიკური სამართლის დოქტორის — ლუკას ვატცენროდეს (Lucas Watzenrode, 1447-1512) გერმანულ ოჯახში. წერდა მხოლოდ გერმანულ და ლათინურ ენებზე და მისი ხელით პოლონურ ენაზე დაწერილი რაიმე დოკუმენტი არ არსებობს.

თუ პრეს-რელიზს დავუბრუნდებით, უნდა ითქვას, რომ, მისი ავტორების აზრით, ამერიკის ფიზიკური საზოგადოების PRL (Physical Review Letters — «წერილები

ფიზიკური მიმოხილვით») ჟურნალში გამოსაქვეყნებლად მომზადებული მათი კვლევის მნიშვნელობა ისაა, რომ მსგავსი ხელოვნური ინტელექტი — ადამიანებზე გაცილებით უფრო ადრე — შეძლებს მომავალში ფიზიკის ახალი კანონების აღმოჩენას, განსაკუთრებით კვანტურ მექანიკაში, რომელიც ახლა მრავალ წინააღმდეგობას აწყდება. სინამდვილეში კი, რაკი ნებისმიერი ასეთი «ხელოვნური ინტელექტი» ადამიანის მიერ შექმნილ მხოლოდ ინსტრუმენტს წარმოადგენს, ეს სავარაუდო აღმოჩენები უფრო ზუსტად იქნებოდა ჩამოყალიბებული სწორედ ადამიანების მიერ — მათივე სრულყოფილი ინსტრუმენტების დახმარებით.

დაბოლოს, გვინდა შევჩერდეთ კიდევ ერთ, საშიშროების შემცველ მაგალითზე, რომლის შესახებ ინფორმაცია 2021 წლის 16 მარტს დაიბეჭდა Guardian-ში.

მკვლევარებმა სტენფორდის უნივერსიტეტიდან შექმნეს ალგორითმი, რომელიც ადამიანის სექსუალურ ორიენტაციას განსაზღვრავს მისი ფოტოგრაფიის საფუძველზე და ამავდროულად მათ მიერ დამტკიცდა, რომ სახის ნაკვთები შეიცავს მნიშვნელოვან ინფორმაციას

ადამიანის არატრადიციული ორიენტაციის შესახებ. კონკრეტულად რა უჩვენა ექსპერიმენტმა? ალგორითმა განსაზღვრა ადამიანების სექსუალური ორიენტაცია ფოტოგრაფიებით გაცნობათა საიტზე 91%-ის სიზუსტით და რთული ეთიკური საკითხები დააყენა. სწორედ ამ მკვლევარებმა აღმოაჩინეს, რომ კომპიუტერულმა ალგორითმა შეძლო გეებისა და ჰეტეროსექსუალური მამაკაცების სწორად გარჩევა შემთხვევათა 81%-ში, ხოლო ქალების — შემთხვევათა 74%-ში. ექსპერიმენტები მიმდინარეობდა რამდენიმე არსებულ ნეიროქსელზე, რომლებიც, ჩვეულებრივ, გამოიყენება პიროვნებათა ამოსაცნობად და მათი სახის ნაკვთების გამოსაკვლევად. ამრიგად გაჩნდა რიგი შეკითხვებისა სექსუალური ორიენტაციის ბიოლოგიური წარმომავლობის შესახებ, სახის ნაკვთების მიხედვით გამოცნობის ტექნოლოგიის ეთიკურობაზე, ასევე მსგავსი პროგრამული უზრუნველყოფის გამოყენების შესაძლებლობაზე, რომელიც არღვევს ადამიანთა კონფიდენციალობას ან ქმნის ამ პროგრამული უზრუნველყოფის ბოროტად გამოყენების შესაძლებლობას არატრადიციული სექსუალური

ორიენტაციის ადამიანების დისკრიმინაციისათვის. კვლევამ უჩვენა, რომ ადამიანის უნარი ამოიცნოს სექსუალური ორიენტაცია მნიშვნელოვნად დაბალია ხელოვნურ ინტელექტთან შედარებით. ადამიანმა შეძლო სწორი გამოცნობა მხოლოდ შემთხვევათა 61 პროცენტში მამაკაცებს შორის და შემთხვევათა 54 პროცენტში ქალებს შორის (ნახ. 17).



ნახ. 17 ადამიანის სექსუალური ორიენტაციის ამსახველი ნიშნების ადგილმდებარეობა მის სახეზე

ექსპერიმენტების მიხედვით, მიღებული შედეგები აძლიერებს იმ თეორიის პოზიციებს, რომელიც ამტკიცებს სექსუალური ორიენტაციის პირდაპირ კავშირს



გარკვეული ჰორმონების ზემოქმედებასთან დაბადებამდე.

სხვანაირად რომ ვთქვათ, ადამიანები უკვე იბადებიან გეებად და ლესბოსელებად.

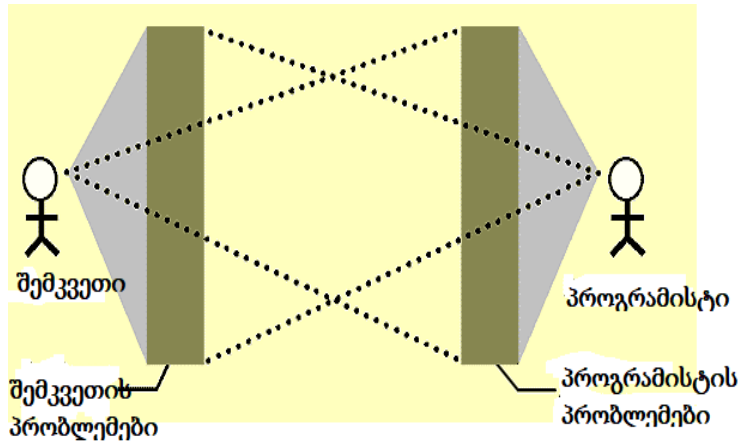
უკიდურესად გაძლიერდა შიში იმის თაობაზე, რომ ტექნოლოგია პერსპექტივაში შეიძლება იყოს გამოყენებული არაჰუმანური მიზნებით, განსაკუთრებით იმ ქვეყნებში, სადაც სექსუალურ უმცირესობათა წარმომადგენლები დევნას განიცდიან.

გამორიცხული არ არის ისიც, რომ ხსენებული ალგორითმის გამოყენება შეიძლება მოხდეს ადამიანთა სექსუალური ორიენტაციის გამოსავლენად მათი თანხმობის გარეშე მრავალი სხვა მიზეზითაც და კაცობრიობა კვლავ ფარსმან სპარსის ეპოქაში აღმოჩნდება.

## **IX. ტიუგუს დიპოლი (ორპოლუსა)**

ხელოვნური ინტელექტის სისტემათა შექმნისას ჩნდება დამახინჯება, რომლის წყაროს ზედმიწევნით მარტივი, მაგრამ ძალიან სასარგებლო, მოდელი აღწერს. პირველად იგი ააგო ესტონელმა მათემატიკოსმა ენ ტიუგუმ

(ესტ. Enn Tõugu, 1935-2020). ამ მოდელს «ტიუგუს დიპოლი» («ტიუგუს ორპოლუსა») ეწოდება. იგი გამოიყურება (ნახ. 18)-ის შესაბამისად.

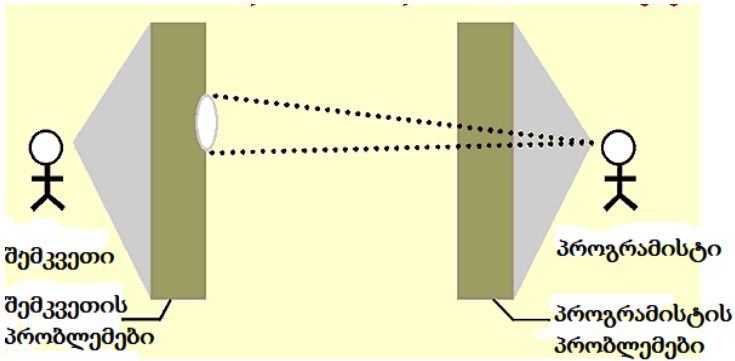


ნახ. 18 ტიუგუს დიპოლი (ორპოლუსა) :  
პირველი ვარიანტი

მოცემულ ვარიანტში დიპოლი გვიჩვენებს, რატომ მახინჯდება «ინტელექტუალური» სისტემის დაპროექტებისას ამ პროცესის ორი მონაწილის აქცია: «დამკვეთის» და «პროგრამისტის». მიზეზი ისაა, რომ თითოეული მათგანი მეორე მონაწილის პრობლემებს აკვირდება

«შორიდან», როცა ეს უცხო, სხვისი სუბსტანცია ძალიან მცირდება და გარკვევით მხოლოდ საკუთარი პრობლემის დეტალები იკვეთება.

დამახინჯებათა კიდევ უფრო ტიპური სიტუაციისათვის ტიუგუს დიპოლის მოდელის სხვა ვარიანტი ნაჩვენებია (ნახ. 19)-ზე. ამ ვარიანტში პროგრამისტი იმდენად ზღუდავს თავისი «გარე» მიმოხილვის არეს, რომ დამკვეთის ყველა პრობლემიდან ხედავს მხოლოდ იმას, რაც ადვილად გაეგება და რისთვისაც თავადაც უნდა პროგრამის დაწერა.



ნახ. 19. ტიუგუს დიპოლი (ორპოლუსა) :  
მეორე ვარიანტი

რა გასაკვირია, რომ ასეთი მიდგომის შედეგად მიღებული «ნაკეთობა» — ინტელექტუალური სისტემა, ინსტრუმენტი — ან უარყოფილი იქნება, ან ხანგრძლივი წამების, წვალების, გაწყალების, ტანჯვის საგანი გახდება.

## X. დასკვნა

ხელოვნური ინტელექტის უკეთესი გაგება, მის «ინტელექტში» ჩაწვდომა, როგორც ეს ბორის ბაროს (ფრ. Boris Barraud) აქვს აღნიშნული თავის ნაშრომში [9], რომლის იდეები ამ ნარკვევშიც არის ფართოდ ასახული, გულისხმობს ამ მეტა და მეგა რევოლუციის (ასევე სხვა მრავალი და ნაირფეროვანი წყაროს) ყველა კოგნიტიური ძალისხმევის გაერთიანებას და საფუძვლიანი შენიშვნების [10, 11] გათვალისწინებასაც. დღეს ხელოვნურ ინტელექტთან ასოცირებულია: უაღრესად მძლავრი კონვოლუციური (ნახვევის ოპერაციასთან დაკავშირებული) ხელოვნური ნეირონული ქსელები, მანქანათა ღრმა სწავლება, ღრუბლოვანი და გაერთიანებული, ანუ ფედერაციული გამოთვლები (ინგლ. *Federated Computing*), მონაცემთა უზარმაზარი მოცულობები...

ეს ისეთი ამოცანების გადაჭრის საშუალებას იძლევა, როგორცაა: სახეთა ამოცნობა, ფოტოგრაფიათა ინტერპრეტაცია, მანქანური თარგმანი, ტექსტების ანალიზი, Data Mining (მონაცემთა მოპოვება, მათი სიღრმისეული ანალიზი), უპილოტო მართვა,, რობოტები, ადამიანების ჩანაცვლება საწარმოებსა და ოფისებში ადამიანისთვის არახელსაყრელი პირობებით.

სამყარომ ფერი იცვალა და ახლა ადამიანის როლი სხვანაირი ხდება. თუ უდიდესი ამერიკელი ფილოსოფოსის ელვინ ტოფლერის (ინგლ. Alvin Toffler, 1928—2016) ცნობილი წიგნის «მესამე ტალღა» (The Third Wave, 1980) დედაარსიდან გამოვალთ, ამ საუკუნეში გაუნათლებელი არა წერა-კითხვის უცოდინარია, როგორც ილია ჭავჭავაძის ეპოქაში, არამედ ის, ვისაც არ შეუძლია სწავლა, ნასწავლის გარდასახვა ან/და ხელახლა სწავლა. არ უნდა შევუშინდეთ რობოტებს და ე.წ. «ხელოვნურ» ინტელექტს! ცნობილი ამერიკელი პროგრამისტი ლარი ტესლერი (Lawrence Gordon Tesler; 1945—2020) იმასაც კი ამბობს, რომ *ხელოვნური ინტელექტი — ეს ისაა, რაც ჯერ კიდევ არ ყოფილა გაკეთებული.*

ცივილიზაციის გადარჩენა ადამიანის მიერ შექმნილ ინტელექტუალურ ინსტრუმენტებში ისახება, მაგრამ ეს სრულებითაც არ ნიშნავს, რომ ევოლუციის სპირალის შემდეგი ხვია ცალსახად არის გონიერი ადამიანის (Homo Sapiens) და ხელოვნური ინტელექტის (Artificial Intelligence) პირმშოს — გონიერი რობოტის (Robo Sapiens) კონვერგენცია (ინტეგრაცია, რაიმე ნიშან-თვისებათა დამთხვევა ერთმანეთისაგან დამოუკიდებელ მოვლენებში), რასაც მრავალი მკვლევარიც აღიარებს.

## XI. ლიტერატურა

[1]. Pope Francis: «Let yourselves be amazed by message of Christmas», ROME REPORTS :

<https://www.romereports.com/en/2020/12/23/pope-francis-let-yourselves-de-amazed-by-the-message-of-christmas/>

[2]. B. Stiegler, La technique et le temps – La faute d'Épiméthée, Galilée, 1994.

[3]. A. Leroi-Gourhan, Le geste et la parole, Albin Michel, 1964.

[4]. G. Perec, Penser / Classer, Hachette, 1985.

[5]. Р. Фейнман, Р. Лейтон, М. Сэндс, Фейнмановские лекции по физике. Том 4. Кинетика. Теплота. Звук. Глава 46. Храповик и собачка. Стр. 138 – 151 :

<http://www.t-z-n.ru/archives/tom4.pdf>

[6]. TLFi : Trésor de la langue Française informatisé, ATILF — CNRS & Université de Lorraine :

<http://atilf.atilf.fr/tlf.htm>

[7]. L. Galiana, «Les conséquences économiques de l'intelligence artificielle», Idées économiques et sociales, 2018, n°192, p. 27.

[8]. A. Turing, «Computing Machinery and Intelligence»,  
Mind : A Quarterly Review of Psychology and Philosophy,  
Volume LIX, Issue 236, October 1950, p. 433–460.

[9]. Sous la direction de Boris Barraud «L'intelligence  
artificielle: Dans toutes ses dimensions», L'Harmattan, 2020,  
308 pages.

[10]. Luc Julia «L'intelligence artificielle n'existe pas», Kindle  
Edition, 2019, 200 pages.

[11]. Giovanni Bruner «No, Artificial Intelligence doesn't  
exist (yet)», Towards Data Science, 12.28.2020 :

[https://towardsdatascience.com/no-artificial-intelligence-  
doesnt-exist-yet-3318d83fdfe8](https://towardsdatascience.com/no-artificial-intelligence-doesnt-exist-yet-3318d83fdfe8)





**გამომცემლობა „უნივერსალი“**

---

თბილისი, 0186, ა. ჯორჯიაძის ქ. №4. ☎: 5(99) 17 22 30; 5(99) 33 52 02  
E-mail: [universal505@ymail.com](mailto:universal505@ymail.com); [gamomcemlobauniversal@gmail.com](mailto:gamomcemlobauniversal@gmail.com)



არჩილ ფრანგიშვილი — ტექნიკის მეცნიერებათა დოქტორი, პროფესორი, საქართველოს მეცნიერებათა ეროვნული აკადემიის ნამდვილი წევრი (აკადემიკოსი), საქართველოს საინჟინრო აკადემიის პრეზიდენტი, რიგი საერთაშორისო და საზღვარგარეთის აკადემიის აკადემიკოსი, საქართველოს ტექნიკური უნივერსიტეტის სამეთვალყურეო საბჭოს პრეზიდენტი, ინფორმატიკისა და გამოთვლითი ტექნიკის, კონფლიქტოლოგიის, მართვის პროცესებისა და სისტემების თეორიის ცნობილი სპეციალისტი.



ოლეგ ნამიჩეიშვილი — ტექნიკის მეცნიერებათა დოქტორი, პროფესორი, საქართველოს და საერთაშორისო საინჟინრო აკადემიათა ნამდვილი წევრი, საქართველოს საბუნებისმეტყველო მეცნიერებათა აკადემიის ნამდვილი წევრი, ინფორმაციული სისტემების, გამოთვლითი და ელექტრონული ტექნიკის, ხელოვნური ნეირონული ქსელების, ფიზიკური მოვლენების კომპიუტერული მოდელირებისა და საიმედოობის მათემატიკური თეორიის ცნობილი სპეციალისტი.



ჟუჟუნა გოგიაშვილი — ფიზიკა-მათემატიკის მეცნიერებათა კანდიდატი, საქართველოს ტექნიკური უნივერსიტეტის ასოცირებული პროფესორი, ფართო მეცნიერული ინტერესებით საიმედოობის მათემატიკური თეორიის, ხელოვნური ნეირონული ქსელების, ფიზიკურ მოვლენათა კომპიუტერული მოდელირების, ინფორმაციული ტექნოლოგიების საინჟინრო ეთიკისა და ხელოვნური ინტელექტის სფეროებში.

ISBN 978-9941-33-418-4

