

როგორ ამოვიცნოთ ხელოვნური ინტელექტით (AI) გენერირებული შინაარსი და გავერკვეთ მასში?

ხელმისაწვდომი და მარტივი მოხმარების AI მოდელები სწავლასა და სხვადასხვა შინაარსის შექმნაში შეიძლება დაგვეხმაროს, თუმცა მათ ასევე შეუძლიათ გააძლიერონ ის რისკები, რომლებსაც **დეზინფორმაცია და მცდარი ინფორმაცია წარმოშობს** ღია საზოგადოებისა და დემოკრატიული დისკურსისთვის. მნიშვნელოვანია, ჩვენი საზიარო ინფორმაციული სივრცეები ხელოვნური ინტელექტის მიერ შექმნილი და ციფრულად შეცვლილი დეზინფორმაციითა და მცდარი ინფორმაციით არ იყოს გადატვირთული.

პრობლემის გადაწყვეტის გზები შეგვიძლია ისეთ თანამედროვე ტექნოლოგიებში ვეძებოთ, როგორიცაა შინაარსის წარმოშობის ინიციატივები (content provenance) და აღმოჩენის პროგრამული უზრუნველყოფა. თუმცა, **აღნიშულ ტექნოლოგიურ ინიციატივებს ბევრი ნაკლი აქვთ; ჩვენ გვჭირდება დამოუკიდებელი ფაქტების გადამოწმების ორგანიზაციები, რათა საზოგადოებას გადამოწმებული ფაქტების ერთიანი ნაკადი მივაწოდოთ.**

წარმოგიდგინოთ მოკლე მიმოხილვას, თუ რას საქმიანობენ პროფესიონალი, დამოუკიდებელი ფაქტების გადამოწმების ორგანიზაციები ხელოვნური ინტელექტის მიერ შექმნილი მცდარი ინფორმაციისა და დეზინფორმაციის იდენტიფიცირებისა და გაქარწყლების კუთხით და რისი სწავლა შეგვიძლია მათგან.

AI-თ შექმნილი შინაარსი სულ უფრო იზრდება

დღესდღეობით, ხელოვნური ინტელექტის მიერ შექმნილი მცდარი ინფორმაციისა და დეზინფორმაციის გამოვლენა დამოუკიდებელი ფაქტების გადამოწმების ორგანიზაციების საქმიანობის მხოლოდ მცირე ნაწილს წარმოადგენს. ციფრულად შეცვლილ შინაარსს ფაქტების გადამოწმების ორგანიზაციები შედარებით მეტ დროს უთმობენ.

თუმცა: EFCSN-ის წევრი ფაქტების გადამოწმების ორგანიზაციების უმეტესობა თანმხდება, რომ ხელოვნური ინტელექტის მიერ გენერირებული და ციფრულად შეცვლილი კონტენტის მნიშვნელობა მომავალში მხოლოდ გაიზრდება. ევროპული არჩევნების კონტექსტში გამოვლენილი ბოლოდროინდელი [მაგალიტები](#) ამ პროგნოზს კიდევ უფრო ამყარებს.

გაითვალისწინეთ: ციფრულად შეცვლილია ნებისმიერი შინაარსი, რომელმაც საგრძნობი ცვლილება განიცადა სხვადასხვა ფორმით, მათ შორის AI ხელსაწყოების გამოყენებით, რათა მანიპულაციურად წარმოჩენილიყო ან შეცვლილიყო ის გზავნილი, რასაც იგი თავდაპირველად ავრცელებდა. ეს არ მოიაზრებს რედაქტირებას სიცხადის ან ხარისხის გაუმჯობესებისთვის.

ხელოვნური ინტელექტით (AI) გენერირებულია ნებისმიერი შინაარსი, რომელიც შეიქმნა ხელოვნური ინტელექტის სისტემის მიერ.

ტექნოლოგიები სწრაფად ვითარდება, თუმცა მხოლოდ მათ იმედად ვერ ვიქნებით

ხელოვნური ინტელექტის ექსპერტები და ფაქტების გადამმონებლები თანხმდებიან: **არსებული ტექნოლოგიური ინსტრუმენტები არ არის საკმარისი AI-გენერირებული ან ციფრულად შეცვლილი შინაარსის გამოსავლენად და გადასამონებლად.**

AI დეტექტორის გამოყენებამდე ექსპერტები გვიჩვენებენ, კარგად გავერკვეთ მათი მუშაობის პრინციპში. ცოტაოდენი სტატისტიკის დახმარებითა და იმის გააზრებით, თუ როგორ სწავლობენ და მუშაობენ AI მოდელები, ფაქტების გადამმონებლებს შეუძლიათ, დაიწყონ AI დეტექტორების ძლიერი და სუსტი მხარეების და მისი წარმატების ალბათობის განსაზღვრა. **მიუხედავად ამისა, ხელსაწყოები შეიძლება მაინც კარგი საწყისი იყოს გადამმონებლის პროცესში.**

შინაარსის წარმოშობა: ისეთი ინიციატივები, როგორცაა **C2PA Specifications**, შეიძლება დაგვეხმაროს მედიაშინაარსის წყაროსა და ისტორიის გადამმონებებაში, თუმცა წვირნიშანი (**watermark**) და ვერიფიკაცია სანდოობის გარანტი ყოველთვის ვერ იქნება.

რა გავლენას ახდენს AI დეზინფორმაცია ადამიანებზე?

“რამდენჯერაც საკუთარი შინაგანი ხმის კარნახით რეაგირებთ, შეგვიძლია ვთქვათ, რომ ეს თქვენს რაციონალურ განსჯას სცილდება.”

- ქრისტინ დუგოანი*

მიზნები: რატომ შეიძლება იყენებდეს მანვ აქტორი AI-ს დეზინფორმაციის შესაქმნელად ან გასავრცელებლად? რა გავლენა შეიძლება ჰქონდეთ მათ რეალურ სამყაროზე?

- სხვა ქვეყანაში ან საზოგადოებაში გავლენის გავრცელება?
- გამოამკარავების თავიდან აცილება და ფაქტების გადამმონებლების დაბნევა?
- საზოგადოებრივ აზრზე ან რწმენაზე არაავთენტური ანგარიშების ქსელების მეშვეობით გავლენის მოხდენა?

ფსიქოლოგია: გავლენის ოპერაციები ხშირად ფსიქოლოგიური მიკერძოებით მანიპულირებას ისახავს მიზნად.

საკუთარი და თქვენი აუდიტორიის გაგება დეზინფორმაციის წინააღმდეგ ბრძოლაში დაგვეხმარება.

* ქრისტინ დუგოანი ინფორმაციული გავლენის მკვლევარია სორბონის უნივერსიტეტში.



გადამონმება მრავალმხრივ მიდგომას და დეტალურ ანალიზს მოითხოვს

რა უნდა მოვიმოქმედოთ, თუ AI-ს გამოვლენის ხელსაწყოები არ მუშაობს? მნიშვნელოვანია, გავიგოთ მტკიცების როგორც კონტექსტი, ისე მისი შინაარსი. ფაქტების გადამმონმებლები ამ მხრივ ექსპერტულ უნარებს ფლობენ. გთავაზობთ რამოდენიმე რჩევას.

“გამოვლენის ხელსაწყოები 100%-იანი სიზუსტით არ მუშაობს და ვერც ვერასდროს იმუშავებს - ჰენკ ვან ესსი**



ყურადღება მიაქციე წყაროს: შეგიძლიათ წყაროს ვინაობის დადასტურება? რაზე საუბრობს და რას აზიარებს იგი? ვინ ენდობა მის მასალებს? რა გავლენა შეიძლება ჰქონდეს ამ შინაარსს მკითხველზე?



დაადგინე სანდოობა: დამოუკიდებლად გადამმონმეთ ინფორმაცია ისეთ სანდო წყაროებთან, როგორიცაა ამ სფეროში პრაქტიკული გამოცდილების მქონე ექსპერტები. დაფიქრდით, აქვს თუ არა ლოგიკური საფუძველი იმას, რასაც ხედავთ?



გამოიყენეთ **ციფრული საგამომძიებლო** ტექნიკები ტრადიციული საგამომძიებო და დოკუმენტური კვლევის გასამყარებლად. ასეთი ტექნიკები შეიძლება იყოს: მონაცემთა შეგროვება (scraping), გეოლოკაცია, ბიომეტრიული ამოცნობა, პატერნების ანალიზი და სხვა.



ისწავლე & ადაპტირდი: AI-თ გენერირებული დეზინფორმაციის შემქმნელები მუდმივად ადაპტირდებიან. თქვენც შეცვალეთ საკუთარი მიდგომა ცვალებად გარემოებებთან მიმართებით.

გააზიარე შენი ნაშრომი

გადამონმებულ მტკიცებასთან ერთად, ექსპერტები გვირჩევენ, გამოვაქვეყნოთ ჩვენს მიერ ჩატარებული ანალიზი გამოყენებული წყაროების ბმულებთან ერთად. ეს შეიძლება დაეხმაროს მკითხველს, მიჰყვეს თქვენს გამოძიებასა და სიღრმისეულად გაიგოს იგი. ზოგიერთ შემთხვევაში, გამოძიების პროცესი უფრო მნიშვნელოვანია, ვიდრე უბრალოდ იმის მტკიცება, რომ შინაარსი AI-ით არის შექმნილი.

** ჰენკ ვან ესსი ციფრული გამოძიებებისა და ფაქტების გადამმონმების ტექნიკების ექსპერტია.

მარტივი გზამკვლევი: სწორი/არასწორი მიდგომები და მინიშნებები

ქვემოთ მოცემულია მინიშნებები, რომლებიც შეიძლება მიუთითებდეს, რომ შინაარსის ნაწილი ხელოვნური ინტელექტის გენერირებული ან ციფრულად შეცვლილია. ამ გზამკვლევაში ნახსენებ სხვა რჩევებთან ერთად (კონტექსტი, საგამოდიებო ტექნიკა და გამოვლენის ხელსაწყოები), ისინი დაგეხმარებათ ამოიცნოთ სიმართლე იმის მიღმა, რასაც ხედავთ.

ტექსტი

- ხშირად (მაგრამ არა ყოველთვის) უკეთ იცავს გრამატიკის წესებს, ვიდრე ადამიანი.
- იყენებს **ზედმეტად ფორმალურ ან სტრუქტურირებულ ენას**, განსაკუთრებით სოციალური მედიის კონტექსტში.
- იყენებს **ზედმეტად ბევრ ზმნიზედებსა და ზედსართავებს**.
- ახასიათებს ადამიანური ემოციების, იუმორის, სარკაზმისა და იდიომების ნაკლებობა.
- შეიძლება **ტოვებდეს კონკრეტულ დეტალებს** (სახელებს, თარიღებს, ლოკაციებს) ან ორიგინალურ იდეებს.
- რაც მთავარია: ტექსტში სწორი ფაქტებია წარმოდგენილი?

ვიდეო

- არ გამოიყენოთ ხელოვნური ინტელექტით გენერირებული გამოსახულების დეტექტორი ვიდეოს კადრებზე.
- **დააკვირდით სახის გამომეტყველებასა და მიმიკებს**, როგორიცაა თვალების ხამხამი, ასევე ემთხვევა თუ არა პირის მოძრაობა აუდიოს.
- შეიძლება ახასიათებდეს **მკვეთრი გადასვლები**.

აუდიო

- შეადარეთ **საეჭვო აუდიო ავთენტიკურ ნიმუშს** ისეთი ინსტრუმენტების გამოყენებით, რომლებსაც შეუძლიათ გამოავლინონ განსხვავებები მეტყველებისა და სუნთქვის რიტმში, ინტონაციაში...
- AI დეტექტორების გამოყენებისას მოერიდეთ დაბალი ხარისხის აუდიო ნიმუშებს...
- შეიძლება გამოვლინდეს **არაბუნებრივი ან მექანიკური მეტყველების ფორმები**, ბუნებრივი პაუზებისა და სუნთქვის ტემპის ნაკლებობით.

ფოტოები

- მოძებნეთ ადგილები **არაბუნებრივი დეტალებით**: სრულყოფილი კანი, ბუნდოვანი ფონი, არაბუნებრივი სილამაზე ან სინათლე და უცნაურობები, როგორიცაა დამატებითი თითები...
- **მოძებნეთ ფოტოგენერატორების ჭვირნიშანი (watermark)**.
- ყურადღება მიაქციეთ გამოსახულებას: ლოგიკურია? აზრობრივია?
- AI დეტექტორების გამოყენებისას აირჩიეთ მაღალი გარჩევადობის ფოტო და მისი ყველაზე ადრე გამოქვეყნებული ვერსია.

