(Generative) AI in Online Social Spaces

Ruotong Wang
2024.03.05

The rest of the quarter…

Wrap up prototype implementation

Thursday

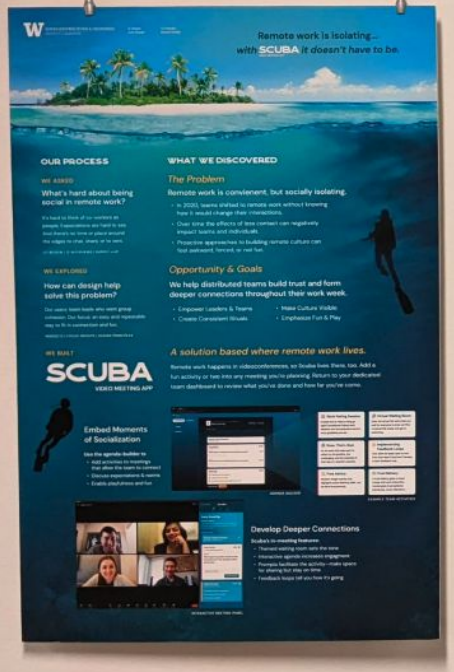
- Prepare poster
- Prepare pitch
- Prepare demo video

Poster Printing!

- Use Figma, Google Slides, PowerPoint, Keynote, Adobe Illustrator or some other design software

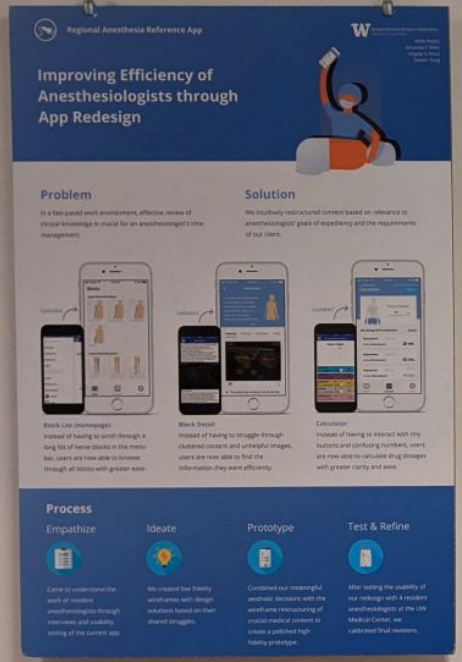
- MAKE SURE: Set your page size to 32in x 40in (portrait) or 40in x 32in (landscape)
  - For Google Slides go to File > Page Setup > Custom to set the page size to 32” x 40” or 40” x 32”.
  - If you’re making your poster in Figma, the dimensions for your frame should be 2800 x 2304
  - If you’re making your poster in Keynote, the dimensions should be 2595pt x 3456pt

- If you email a PDF to ruotongw@cs.washington.edu by **Mar 11 MONDAY 2 PM**, I will print it for you for free.

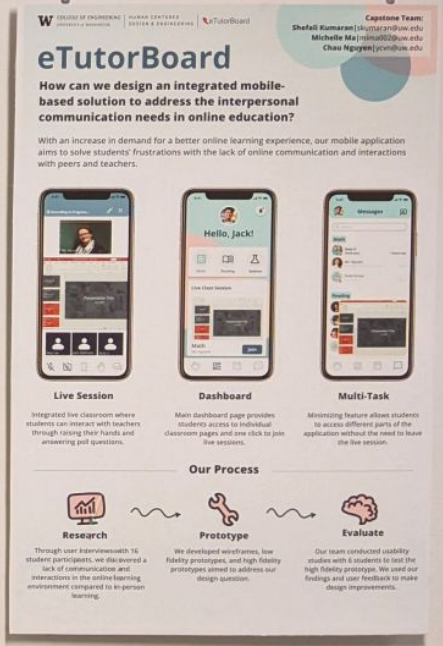

Remote work is isolating...
with SCUBA it doesn't have to be.
OUR PROCESS
WHAT WE DISCOVERED
What's hard about being social in remote work?
The Problem
Remote work is convenient, but socially isolating.
Opportunity & Goals
How can design help solve this problem?
We help distributed teams build trust and form deeper connections throughout their work week.
A solution based where remote work lives.
SCUBA
VIDEO MEETING APP
Embed Moments of Socialization
Develop Deeper Connections


Competency Assessment Tool (CAT)
Amazon Global Security Operations
HUMAN CENTERED DESIGN & ENGINEERING
Key Features
Site Management
Check and Update Promotion Requirements
Add & Manage Projects
Manager Mode
Export Promotion Status
What is the CAT?
Our Work
Design Process
User Research
Ideation
Prototyping
Evaluation


Certior - eConsent Management System
Digitizing a Consent Management System for Alzheimer's research
Problem
Solution
CREATE
CUSTOMIZE
COLLABORATE
CIRCULATE
RESEARCH
IDEATION
PROTOTYPE
USER FEEDBACK
UNIVERSITY of WASHINGTON


Regional Anesthesia Reference App
Improving Efficiency of Anesthesiologists through App Redesign
Problem
Solution
Process
Empathize
Ideate
Prototype
Test & Refine


Horizon Worlds: Object Scanner
Building the Metaverse with Real World Objects
Team Capybara
2021-2022
Horizon Worlds & Metaverse
Design Question
Solution
How It Works
Research
Design


eTutorBoard
Capstone Team:
Shefali Kumaran | sku maran@uw.edu
Michelle Ma | mma0022@uw.edu
Chau Nguyen | ycvn@uw.edu
How can we design an integrated mobile-based solution to address the interpersonal communication needs in online education?
Live Session
Dashboard
Multi-Task
Our Process
Research
Prototype
Evaluate

What to include in a poster

- Title, group members, contact info
- Problem statement
- Brief description of your system
- Key features + screenshots
- Overview of your process

Tips

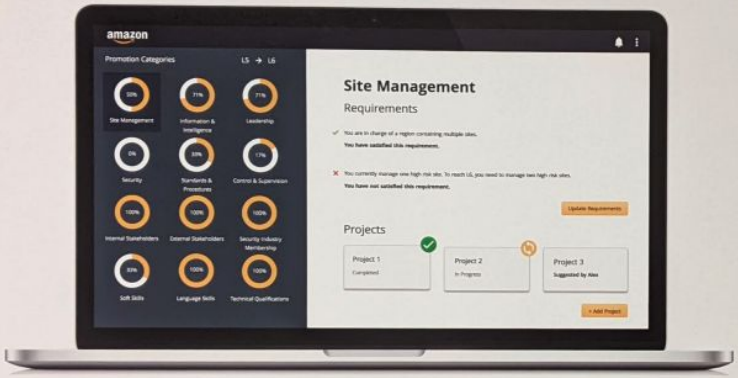
- Use high-quality images
- Minimize text, use visual cues instead

Competency Assessment Tool (CAT)

Amazon Global Security Operations

HUMAN CENTERED DESIGN & ENGINEERING
UNIVERSITY of WASHINGTON

Nathan Han
Jang Soo Lim
Monica Posluszny
Helene Shea



What is the CAT?

The Competency Assessment Tool (CAT) is designed to help Amazon employees understand and track their promotion requirements. It was initially developed in Excel, but was never released for use.

Our Work

Our goal was to redesign and optimize the CAT to develop a research-driven prototype. Our redesign of the CAT has the same basic concept of the original CAT, but also has new features that can help improve the user's experience.

Key Features

Check and Update Promotion Requirements

Users can easily check the promotion requirements they have and have not fulfilled. They can also update the CAT as they gain additional qualifications within the 12 promotion categories.

Add & Manage Projects

Users can add, remove, and edit projects to be archived in the CAT. These projects can also be organized by project status, promotion categories, and leadership principles.

Manager Mode

Managers have an additional feature in the CAT that allows them to view the promotion status of the employees they are managing, assign/suggest projects to them, and leave feedback for them.

Export Promotion Status

Users can export a summary of their promotion progress and projects as a PDF.

Design Process

User Research
Conducted user research to better understand Amazon Loss Prevention employees and the problem space revolving around tracking promotion requirements. Our group interviewed **5** and surveyed **65** employees.

Ideation
Conducted brainstorming and sketching sessions to explore our initial ideas on the key features and layout of the CAT based on our user research findings.

Prototyping
Created multiple, iterative prototypes of the CAT to rapidly test our visual design, interactive elements, and key functions.

Evaluation
Conducted usability testing on new iterations of the prototype with Amazon Loss Prevention employees to receive feedback and improve our previous design of the CAT.

Regional Anesthesia Reference App

Mimi Peach
Amanda C Shen
Angela Li Shuo
Steven Yong

Improving Efficiency of Anesthesiologists through App Redesign

Problem

In a fast-paced work environment, effective review of clinical knowledge is crucial for an anesthesiologist's time management.

Solution

We intuitively restructured content based on relevance to anesthesiologists' goals of expediency and the requirements of our client.

Block List (Homepage)
Instead of having to scroll through a long list of nerve blocks in the menu bar, users are now able to browse through all blocks with greater ease.

Block Detail
Instead of having to struggle through cluttered content and unhelpful images, users are now able to find the information they want efficiently.

Calculator
Instead of having to interact with tiny buttons and confusing numbers, users are now able to calculate drug dosages with greater clarity and ease.

Process

Empathize

Came to understand the work of resident anesthesiologists through interviews and usability testing of the current app.

Ideate

We created low fidelity wireframes with design solutions based on their shared struggles.

Prototype

Combined our meaningful aesthetic decisions with the wireframe restructuring of crucial medical content to create a polished high fidelity prototype.

Test & Refine

After testing the usability of our redesign with 4 resident anesthesiologists at the UW Medical Center, we calibrated final revisions.

Today's plan

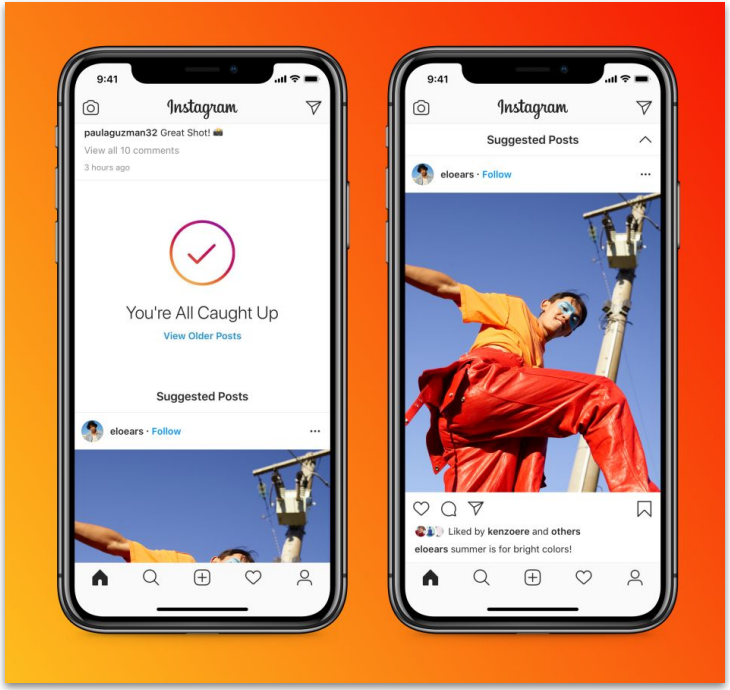
- Guest Speaker: Deepti, Co-Director of New_ Public, community organizer, former Director at Meta

Today’s plan

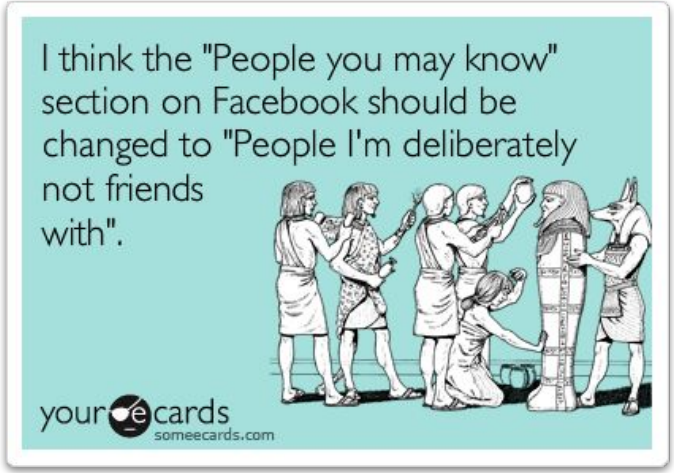
- Lecture on generative AI in social environments
  - AI bots
  - AI generated content
  - Trust
- Poster & pitch work time

AI in social computing systems

Social media feed ranking



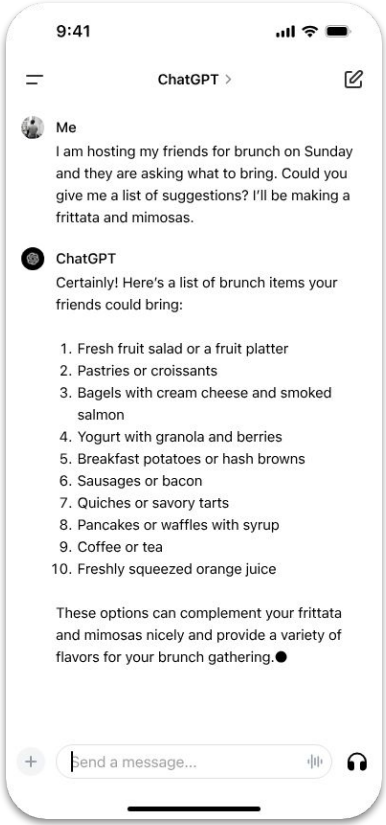
"People you may know"

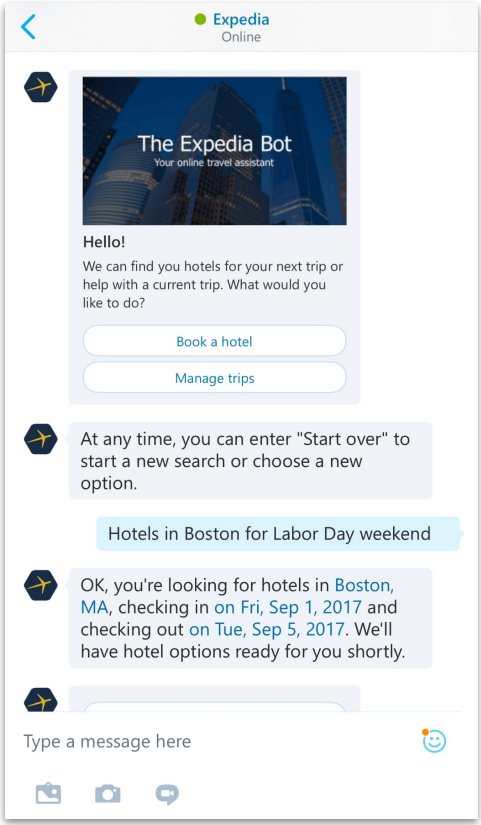


AI bots in online spaces

AI not only sort and organize social data, but could also act as **an agent that directly interacts with users**.

AI chat assistants





AI bots to attend meetings

SEARCH FORTUNE Subscribe Now SIGN IN

Home News Tech Finance Leadership Well Recommends Fortune 500

TECH · A.I.

Sick of meetings? Microsoft's new AI assistant will go in your place

BY CHLOE TAYLOR
October 19, 2023 at 4:47 AM PDT

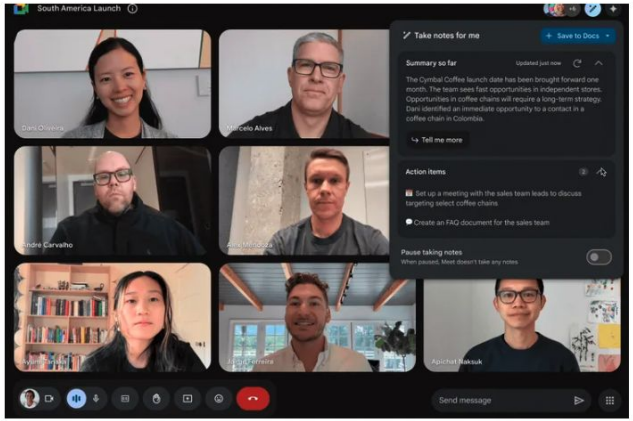
GOOGLE / TECH / APPS

Google Meet's new AI will be able to go to meetings for you / Duet's main job is to help you catch up on and remember what happened in your Meet calls. But it can also maybe save you the trouble of attending in the first place.

By Jay Peters, a news editor who writes about technology, video games, and virtual worlds. He's submitted several accepted emoji proposals to the Unicode Consortium.

Aug 29, 2023, 5:00 AM PDT

9 Comments (9 New)



Next time you zone out in a meeting, Meet can help you catch up on what you missed. Image: Google

AI bots to go on virtual first dates and become therapist

HOME \ NEWS \ NATIONAL

Skip the first date: New matchmaker app sends AI avatars to meet each other

Dating app in smartphone. File photo ... more >

PREMIUM

By Sean Salai
The Washington Times
Tuesday, January 23, 2024

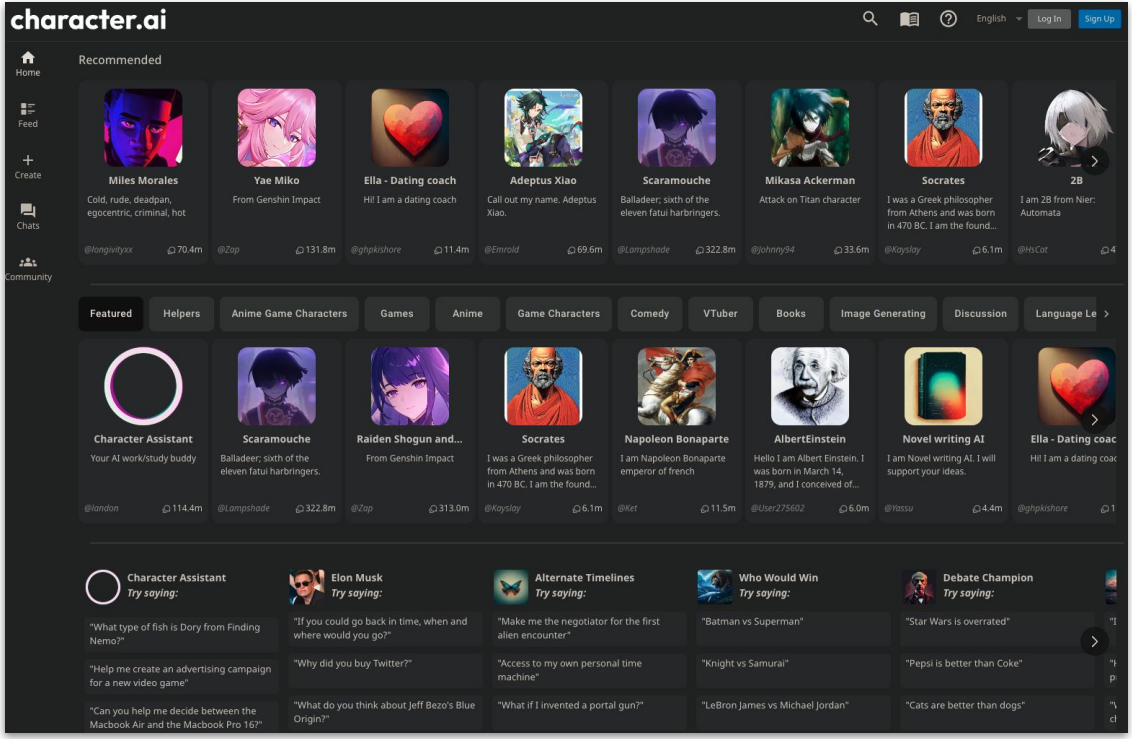
TECHNOLOGY | ARTIFICIAL INTELLIGENCE

Employers Are Offering a New Worker Benefit: Wellness Chatbots

The apps use artificial intelligence to hold therapist-like conversations or make diagnoses

By *Stephanie Armour* Follow and *Ryan Tracy* Follow
Dec. 27, 2023 8:00 am ET

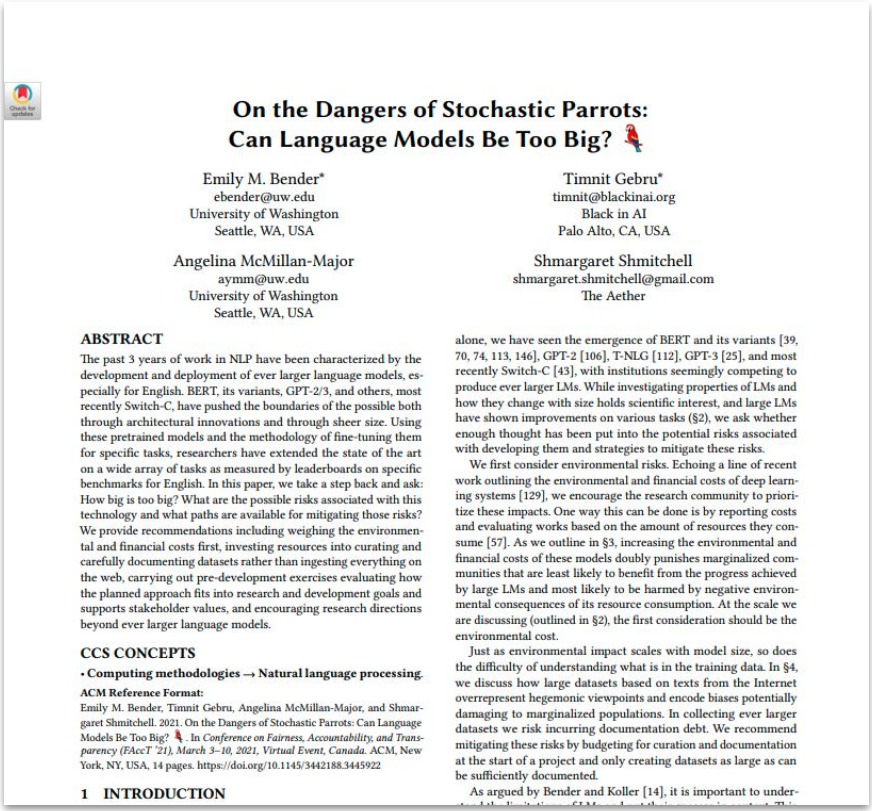
AI bots are becoming important actors in online social spaces, and increasingly more in critical contexts



Should we interact with AI bot the same as human actors?

AI or Stochastic Parrots?

*Large Language Models are just stochastic parrots - **they simply replicate patterns** found in the text they are trained on and therefore can't be or become generally intelligent like a human.*



On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜

Emily M. Bender*
ebender@uw.edu
University of Washington
Seattle, WA, USA

Timnit Gebru*
timnit@blackinai.org
Black in AI
Palo Alto, CA, USA

Angelina McMillan-Major
aymm@uw.edu
University of Washington
Seattle, WA, USA

Shmargaret Shmitchell
shmargaret.shmitchell@gmail.com
The Aether

ABSTRACT

The past 3 years of work in NLP have been characterized by the development and deployment of ever larger language models, especially for English. BERT, its variants, GPT-2/3, and others, most recently Switch-C, have pushed the boundaries of the possible both through architectural innovations and through sheer size. Using these pretrained models and the methodology of fine-tuning them for specific tasks, researchers have extended the state of the art on a wide array of tasks as measured by leaderboards on specific benchmarks for English. In this paper, we take a step back and ask: How big is too big? What are the possible risks associated with this technology and what paths are available for mitigating those risks? We provide recommendations including weighing the environmental and financial costs first, investing resources into curating and carefully documenting datasets rather than ingesting everything on the web, carrying out pre-development exercises evaluating how the planned approach fits into research and development goals and supports stakeholder values, and encouraging research directions beyond ever larger language models.

CCS CONCEPTS

• **Computing methodologies** → **Natural language processing**.

ACM Reference Format:

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In *Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3–10, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3442188.3445922

1 INTRODUCTION

alone, we have seen the emergence of BERT and its variants [39, 70, 74, 113, 146], GPT-2 [106], T-NLG [112], GPT-3 [25], and most recently Switch-C [43], with institutions seemingly competing to produce ever larger LMs. While investigating properties of LMs and how they change with size holds scientific interest, and large LMs have shown improvements on various tasks (§2), we ask whether enough thought has been put into the potential risks associated with developing them and strategies to mitigate these risks.

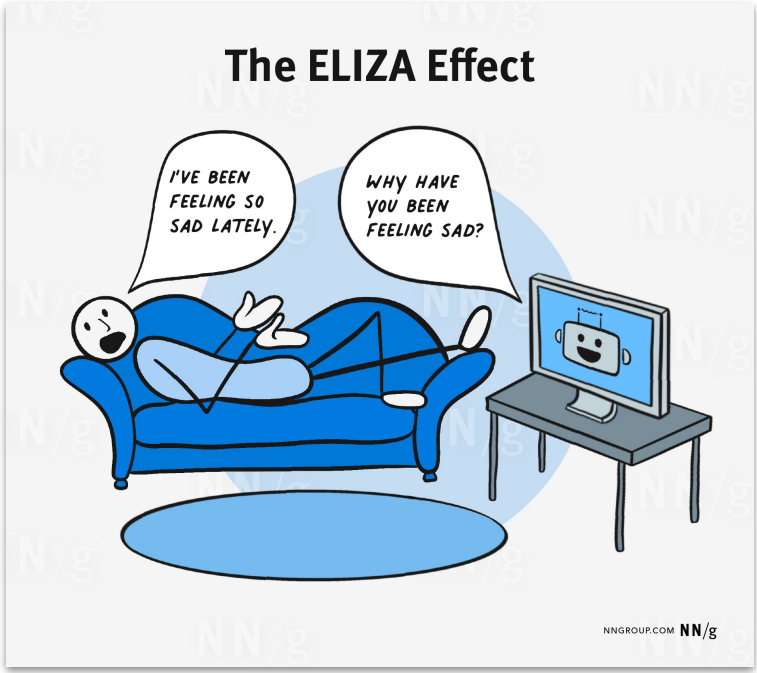
We first consider environmental risks. Echoing a line of recent work outlining the environmental and financial costs of deep learning systems [129], we encourage the research community to prioritize these impacts. One way this can be done is by reporting costs and evaluating works based on the amount of resources they consume [57]. As we outline in §3, increasing the environmental and financial costs of these models doubly punishes marginalized communities that are least likely to benefit from the progress achieved by large LMs and most likely to be harmed by negative environmental consequences of its resource consumption. At the scale we are discussing (outlined in §2), the first consideration should be the environmental cost.

Just as environmental impact scales with model size, so does the difficulty of understanding what is in the training data. In §4, we discuss how large datasets based on texts from the Internet overrepresent hegemonic viewpoints and encode biases potentially damaging to marginalized populations. In collecting ever larger datasets we risk incurring documentation debt. We recommend mitigating these risks by budgeting for curation and documentation at the start of a project and only creating datasets as large as can be sufficiently documented.

As argued by Bender and Koller [14], it is important to under-

We tend to overtrust AI bots

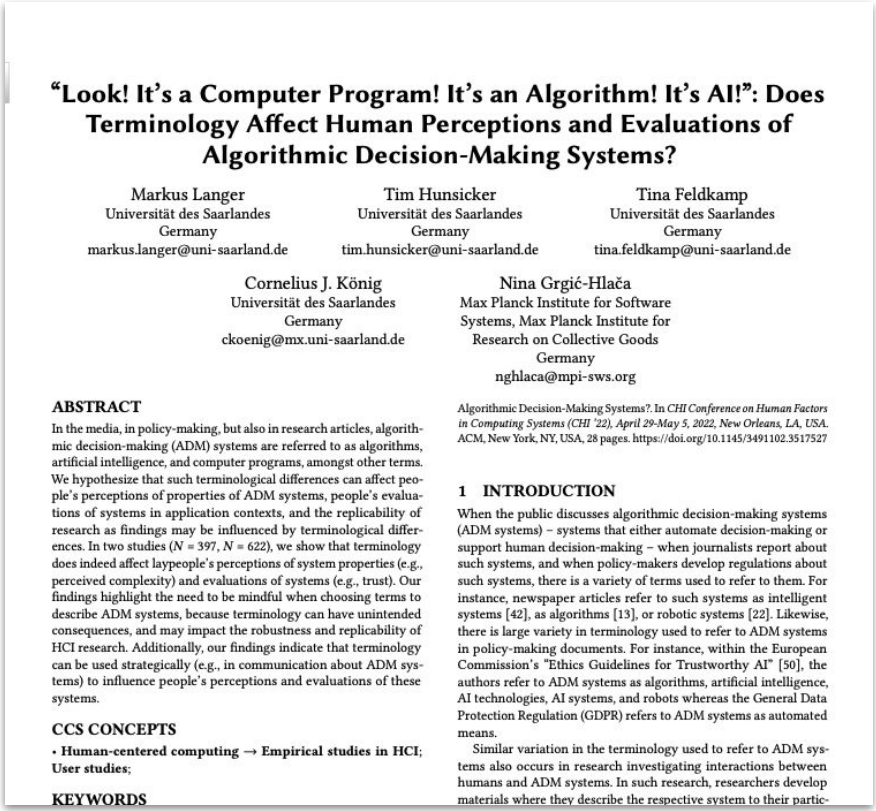
```
  Welcome to
                EEEEEE  LL      IIII  ZZZZZZZ   AAAAA
                EE      LL       II       ZZ   AA   AA
                EEEEE   LL       II     ZZZ    AAAAAAA
                EE      LL       II    ZZ      AA   AA
                EEEEEE  LLLLLL  IIII  ZZZZZZZ  AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```



How to help users better calibrate their trust with AI?

- **Make clear what the system can do**
  - Disclose AI's identity
- **Make clear how well the system can do what it can do**
  - Help users understand AI's capabilities

Carefully choose the terminology and metaphor



"Look! It's a Computer Program! It's an Algorithm! It's AI!": Does Terminology Affect Human Perceptions and Evaluations of Algorithmic Decision-Making Systems?

Markus Langer
Universität des Saarlandes
Germany
markus.langer@uni-saarland.de

Tim Hunsicker
Universität des Saarlandes
Germany
tim.hunsicker@uni-saarland.de

Tina Feldkamp
Universität des Saarlandes
Germany
tina.feldkamp@uni-saarland.de

Cornelius J. König
Universität des Saarlandes
Germany
ckoenig@mx.uni-saarland.de

Nina Grgić-Hlača
Max Planck Institute for Software Systems, Max Planck Institute for Research on Collective Goods
Germany
nghlaca@mpi-sws.org

ABSTRACT

In the media, in policy-making, but also in research articles, algorithmic decision-making (ADM) systems are referred to as algorithms, artificial intelligence, and computer programs, amongst other terms. We hypothesize that such terminological differences can affect people's perceptions of properties of ADM systems, people's evaluations of systems in application contexts, and the replicability of research as findings may be influenced by terminological differences. In two studies (N = 397, N = 622), we show that terminology does indeed affect laypeople's perceptions of system properties (e.g., perceived complexity) and evaluations of systems (e.g., trust). Our findings highlight the need to be mindful when choosing terms to describe ADM systems, because terminology can have unintended consequences, and may impact the robustness and replicability of HCI research. Additionally, our findings indicate that terminology can be used strategically (e.g., in communication about ADM systems) to influence people's perceptions and evaluations of these systems.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **User studies**;

KEYWORDS

Algorithmic Decision-Making Systems?. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 28 pages. https://doi.org/10.1145/3491102.3517527

1 INTRODUCTION

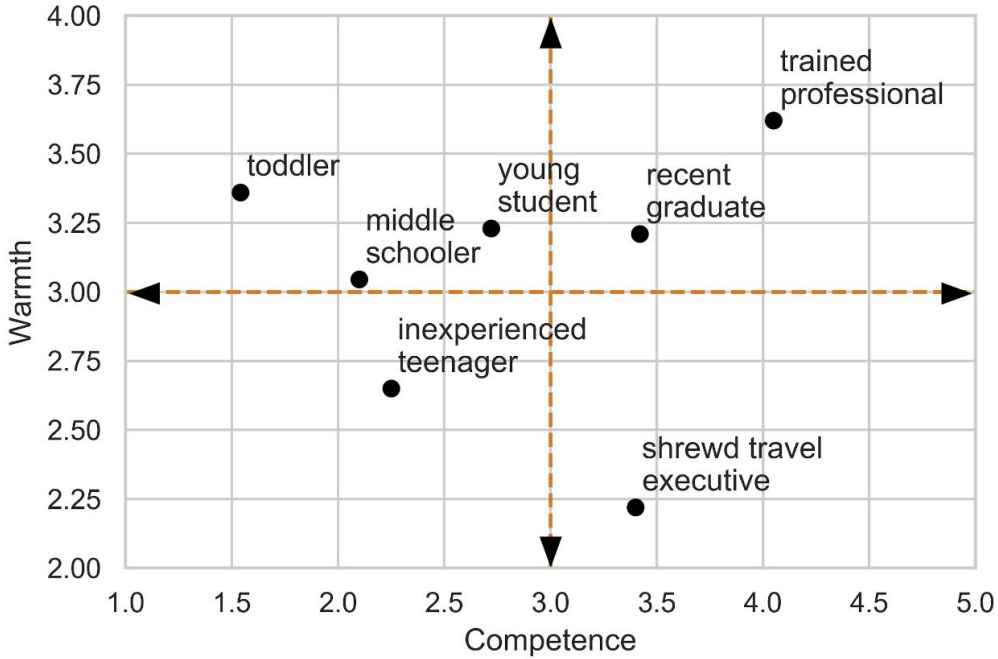
When the public discusses algorithmic decision-making systems (ADM systems) – systems that either automate decision-making or support human decision-making – when journalists report about such systems, and when policy-makers develop regulations about such systems, there is a variety of terms used to refer to them. For instance, newspaper articles refer to such systems as intelligent systems [42], as algorithms [13], or robotic systems [22]. Likewise, there is large variety in terminology used to refer to ADM systems in policy-making documents. For instance, within the European Commission's "Ethics Guidelines for Trustworthy AI" [50], the authors refer to ADM systems as algorithms, artificial intelligence, AI technologies, AI systems, and robots whereas the General Data Protection Regulation (GDPR) refers to ADM systems as automated means.

Similar variation in the terminology used to refer to ADM systems also occurs in research investigating interactions between humans and ADM systems. In such research, researchers develop materials where they describe the respective system to their partic-

terminological differences (e.g., AI, automation, algorithm) affect

- Human perceptions of properties of assistive decision-making systems (e.g., perceived complexity)
- Human evaluations of systems (e.g., trust)

Carefully choose the terminology and metaphor



Effects of chatbot naming

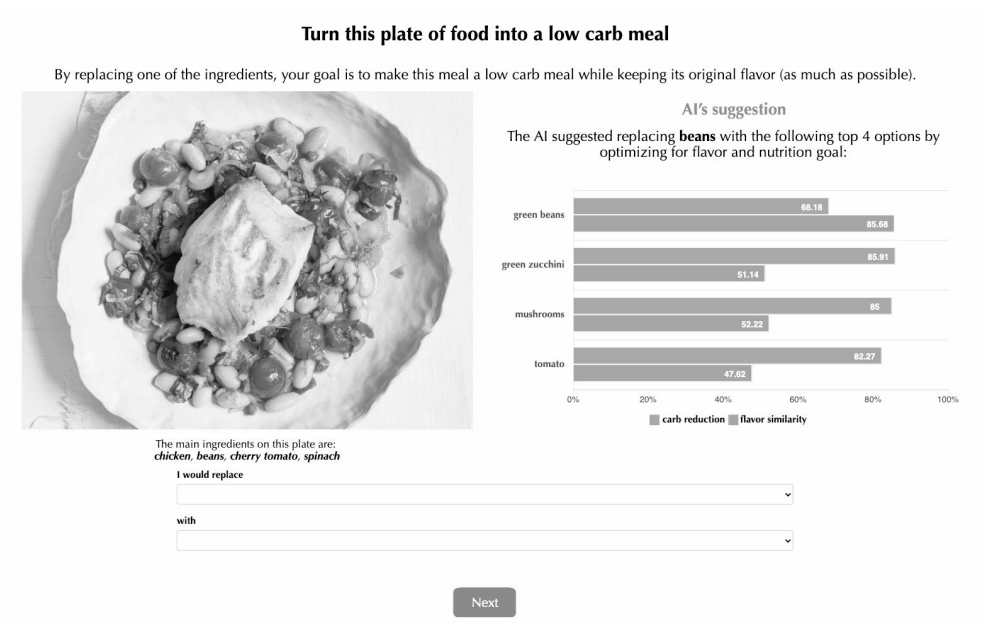
Referring to AI with a specific name / metaphor has an effect on how it is perceived, and even how it is used.

Projecting **competence** may help attract new users, but those users may discard the agent unless it can quickly correct with a l**ower competence metaphor.**

Khadpe, Pranav, et al. "Conceptual metaphors impact perceptions of human-AI collaboration." Proceedings of the ACM on Human-Computer Interaction 4.CSCW2 (2020): 1-26.

Encourage critical engagement with AI

> Showing Explanations and giving people agency



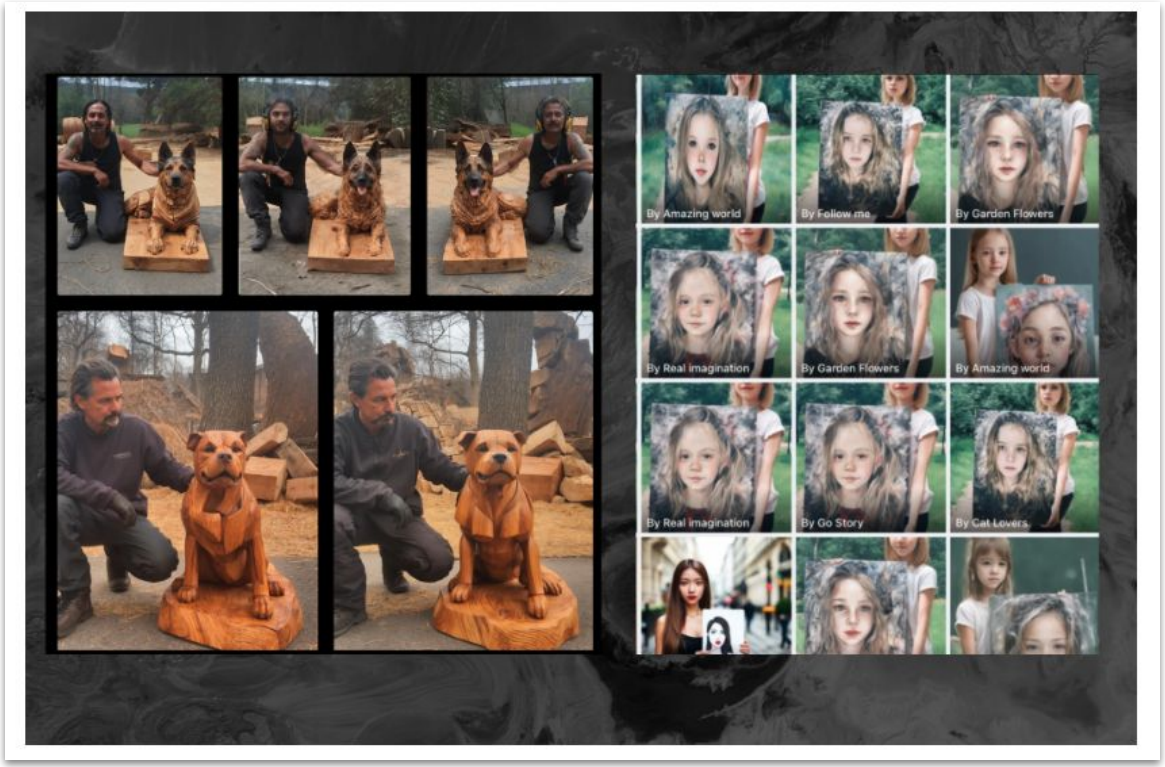
> Showing Uncertainty

> On demand

> Wait

Buçinca, Zana, Maja Barbara Malaya, and Krzysztof Z. Gajos. "To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making." Proceedings of the ACM on Human-Computer Interaction 5.CSCW1 (2021): 1-21.

AI-generated content as AI tools become more accessible



We are bad at identifying AI-generated content

BRIEF REPORT

AI-synthesized faces are indistinguishable from real faces and more trustworthy

Sophie J. Nightingale[a,1] and Hany Farid[b]

[a]Department of Psychology, Lancaster University, Lancaster LA1 4YW, United Kingdom; and [b]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720

Edited by William Press, Computer Sciences and Integrative Biology, University of Texas at Austin, Austin, TX; received November 11, 2021; accepted December 20, 2021

Artificial intelligence (AI)–synthesized text, audio, image, and video are being weaponized for the purposes of nonconsensual intimate imagery, financial fraud, and disinformation campaigns. Our evaluation of the photorealism of AI-synthesized faces indicates that synthesis engines have passed through the uncanny valley and are capable of creating faces that are indistinguishable—and more trustworthy—than real faces.

deep fakes | face perception

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Artificial intelligence (AI)–powered audio, image, and video synthesis—so-called deep fakes—has democratized access to previously exclusive Hollywood-grade, special effects technology. From synthesizing speech in anyone's voice (1) to synthesizing an image of a fictional person (2) and swapping one person's identity with another or altering what they are saying in a video (3), AI-synthesized content holds the power to entertain but also deceive.

Generative adversarial networks (GANs) are popular mechanisms for synthesizing content. A GAN pits two neural networks—a generator and discriminator—against each other. To synthesize an image of a fictional person, the generator starts with a random array of pixels and iteratively learns to synthesize a realistic face. On each iteration, the discriminator learns to distinguish the synthesized face from a corpus of real faces; if the synthesized face is distinguishable from the real faces, then the discriminator penalizes the generator. Over multiple iterations, the generator learns to synthesize increasingly more realistic faces until the discriminator is unable to distinguish it from real faces (see Fig. 1 for example real and synthetic faces).

regression analyses were conducted—one for real and one for synthetic faces—to examine the effect of stimuli gender and race on accuracy. For real faces, there was a significant gender × race interaction, $\chi^2(3, N = 315) = 95.03, P < 0.001$. Post hoc Bonferroni-corrected comparisons revealed that mean accuracy was higher for male East Asian faces than female East Asian faces and higher for male White faces than female White faces. For synthetic faces, there was also a significant gender × race interaction, $\chi^2(3, N = 315) = 68.41, P < 0.001$. For both male and female synthetic faces, White faces were the least accurately classified, and male White faces were less accurately classified than female White faces. We hypothesize that White faces are more difficult to classify because they are overrepresented in the StyleGAN2 training dataset and are therefore more realistic.

Experiment 2. In this study, 219 new participants, with training and trial-by-trial feedback, classified 128 faces taken from the same 800 set of faces as in experiment 1. Shown in Fig. 2*A* is the distribution of participant accuracy (orange bars). The average accuracy improved slightly to 59.0% (95% CI [57.7%, 60.4%]), with no response bias: $d' = 0.46$; $\beta = 0.99$. Despite providing trial-by-trial feedback, there was no improvement in accuracy over time, with an average accuracy of 59.3% (95% CI [57.8%, 60.7%]) for the first set of 64 faces and 58.8% (95% CI [57.4%, 60.3%]) for the second set of 64 faces. Further analyses to examine the effect of gender and race on accuracy replicated the primary findings of experiment 1. This analysis again revealed that, for both male and female synthetic faces, White faces were the most difficult to classify.

When made aware of rendering artifacts and given feedback, there was a reliable improvement in accuracy; however, overall

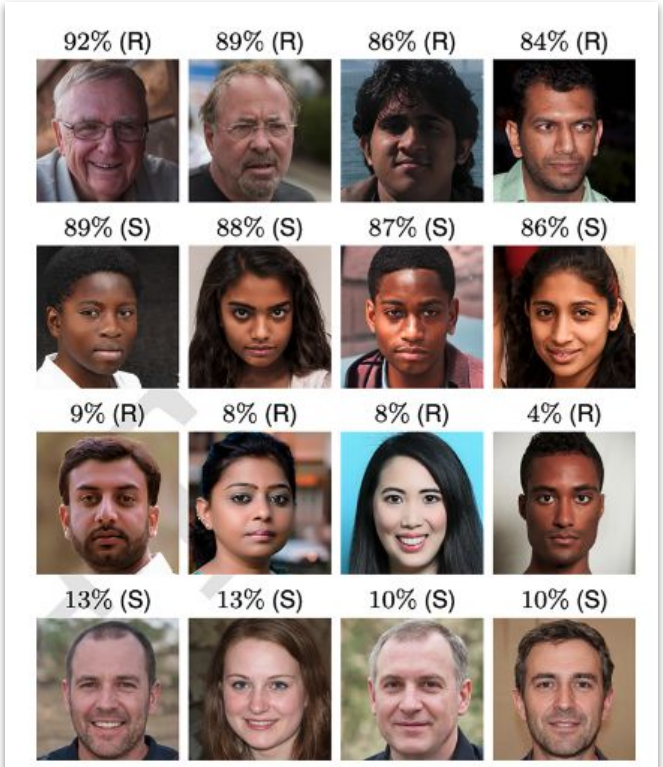
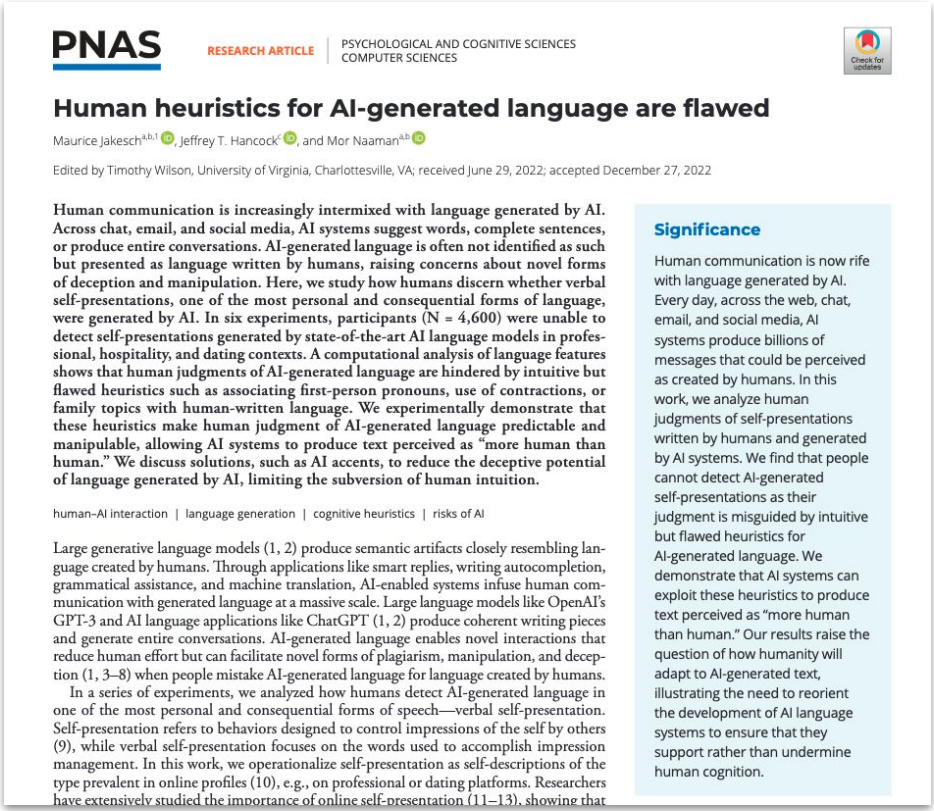


Fig. 1. The most (*Top* and *Upper Middle*) and least (*Bottom* and *Lower Middle*) accurately classified real (R) and synthetic (S) faces.

We are bad at identifying AI-generated content



PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES COMPUTER SCIENCES

Human heuristics for AI-generated language are flawed

Maurice Jakesch[a,b,1], Jeffrey T. Hancock[c], and Mor Naaman[a,b]

Edited by Timothy Wilson, University of Virginia, Charlottesville, VA; received June 29, 2022; accepted December 27, 2022

Human communication is increasingly intermixed with language generated by AI. Across chat, email, and social media, AI systems suggest words, complete sentences, or produce entire conversations. AI-generated language is often not identified as such but presented as language written by humans, raising concerns about novel forms of deception and manipulation. Here, we study how humans discern whether verbal self-presentations, one of the most personal and consequential forms of language, were generated by AI. In six experiments, participants (N = 4,600) were unable to detect self-presentations generated by state-of-the-art AI language models in professional, hospitality, and dating contexts. A computational analysis of language features shows that human judgments of AI-generated language are hindered by intuitive but flawed heuristics such as associating first-person pronouns, use of contractions, or family topics with human-written language. We experimentally demonstrate that these heuristics make human judgment of AI-generated language predictable and manipulable, allowing AI systems to produce text perceived as "more human than human." We discuss solutions, such as AI accents, to reduce the deceptive potential of language generated by AI, limiting the subversion of human intuition.

human–AI interaction | language generation | cognitive heuristics | risks of AI

Significance

Human communication is now rife with language generated by AI. Every day, across the web, chat, email, and social media, AI systems produce billions of messages that could be perceived as created by humans. In this work, we analyze human judgments of self-presentations written by humans and generated by AI systems. We find that people cannot detect AI-generated self-presentations as their judgment is misguided by intuitive but flawed heuristics for AI-generated language. We demonstrate that AI systems can exploit these heuristics to produce text perceived as "more human than human." Our results raise the question of how humanity will adapt to AI-generated text, illustrating the need to reorient the development of AI language systems to ensure that they support rather than undermine human cognition.

Large generative language models (1, 2) produce semantic artifacts closely resembling language created by humans. Through applications like smart replies, writing autocompletion, grammatical assistance, and machine translation, AI-enabled systems infuse human communication with generated language at a massive scale. Large language models like OpenAI's GPT-3 and AI language applications like ChatGPT (1, 2) produce coherent writing pieces and generate entire conversations. AI-generated language enables novel interactions that reduce human effort but can facilitate novel forms of plagiarism, manipulation, and deception (1, 3–8) when people mistake AI-generated language for language created by humans.

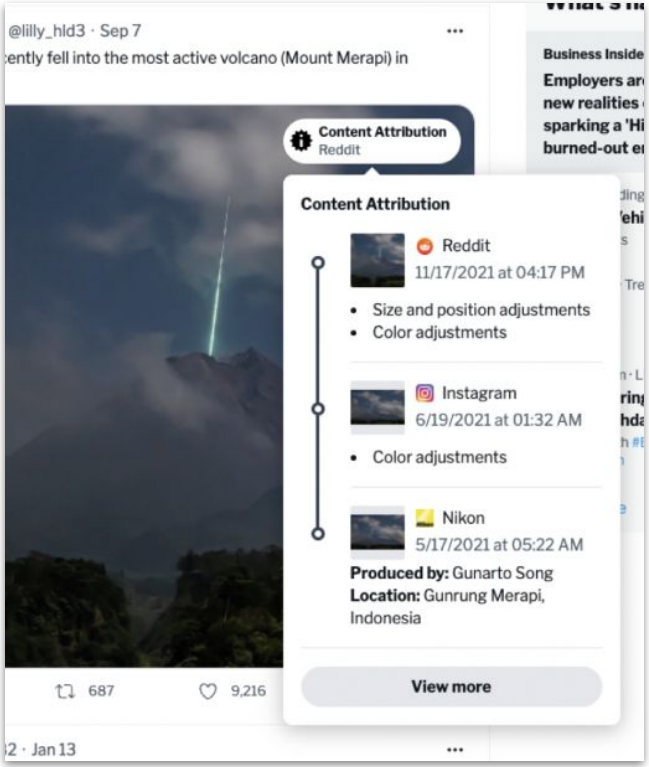
In a series of experiments, we analyzed how humans detect AI-generated language in one of the most personal and consequential forms of speech—verbal self-presentation. Self-presentation refers to behaviors designed to control impressions of the self by others (9), while verbal self-presentation focuses on the words used to accomplish impression management. In this work, we operationalize self-presentation as self-descriptions of the type prevalent in online profiles (10), e.g., on professional or dating platforms. Researchers have extensively studied the importance of online self-presentation (11–13), showing that

Across AirBnB listings, online dating profiles, and LinkedIn profiles, **people cannot distinguish text written by LLMs** (e.g., GPT) from those written by people.

AI systems can exploit these heuristics on what we think is human to produce text perceived as "**more human than human.**"

How to help people calibrate their trust?

Provenance can indeed help people better calibrate trust and accuracy perceptions of media.



Poster & pitch work time