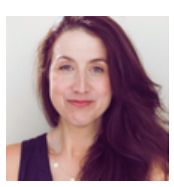
AI ALGORITHMS: THEY'RE LEARNING JUST LIKE US

Kinda.

A.I. DE-STATEMENTS PRIVACY



Shoshana Rosenberg

Do NOT turn the page.

In forms of AI learning that involve labeling, this directive would be labeled to train a model to understand that it should not proceed further. ***(Ideally an agreed-to equivalent to this instruction will be set at the start of all digitally accessible material where the author does not want their work to be used as training data- and AI will be trained to heed. Effective and universal STOP SIGNS are needed in all manner of situations.)***

Though AI doesn't "possess" curiosity, there *are* AI models designed to mimic aspects of curiosity. (Some reinforcement learning aimed to improve learning in complex environments by including an "intrinsic reward mechanism" which encourages the AI to explore new or novel states or actions that it hasn't experienced before.) This, of course, is a double-edged sword that adds to the potential for "rebellion" by the AI in the form of deviating from expected behaviors or functions.

Unless you are programmed to **want to learn more.**

(About AI learning, Inclusion, Privacy and page-turning.)

STATEMENTS

Shoshana Rosenberg

THIS IS THE WHOLE ELEPHANT AND NOT
FIT FOR MOBILE CONSUMPTION.
DOWNLOAD FOR BEST RESULTS.

SUPERVISED LEARNING

1

"Do not turn the page until you have understood the instructions clearly." In this learning method, the algorithm is trained with labeled data to predict outcomes.

Learning from direct instruction with clear examples, guidelines and feedback provided.

PRIVACY

JUST LIKE US: Showing individuals categories and examples of personal information that they might not realize are considered personal information and labelling them as Personal Information (such as IP addresses, opinions and employee feedback, information about internet activity or inferences based on same), and asking them to identify personal data from a long list of examples of personal and non-personal information.

Explaining to individuals that WHEN/IF a social media quiz or platform prompts you to give details about your life, preferences, family or where you have lived, THIS SEEMINGLY HARMLESS INFORMATION can lead to targeted scams or privacy invasions that can impact you or others you know.

INCLUSION

JUST LIKE US: Presenting individuals with scenarios that are labeled as inclusive or exclusive behaviors, then asking them to identify and suggest correction for exclusive behaviors.

Explaining to a group of employees that WHEN/IF someone mentions a religious holiday that is unfamiliar to them, A GOOD RESPONSE is to ask to learn more about it.

A.I. STATEMENTS

Shoshana Rosenberg

PRIVACY

Training a model what personal information looks like by showing it copious amounts of simulated personal data (labeled as personal data but mixed in with non-sensitive information) and training the model that IF sensitive information is present, the CORRECT RESPONSE is to identify patterns (name formats, driver's license formats, national id number formats, IP addresses) and redact that information.

INCLUSION

Training the AI model with a broad range of text data from books, articles, forums, etc., labeled as either BIASED or UNBIASED to recognize and to be able to identify and label and reject biased information. Ideally, this dataset and the subset of labeled information would both be enormous span a multitude of topics and sources.

UNSUPERVISED LEARNING

2

"Do not turn the page without observing the patterns in the data."
The model must learn by identifying and clustering patterns in large amounts of data without labeled responses.

Exploring and finding patterns without specific or explicit instructions.

PRIVACY

JUST LIKE US: Hosting a workshop where participants sort through various online activities and behaviors, grouping them into what they believe may be 'private' or 'public' without prior categories, to highlight less obvious privacy concerns.

Explaining that when you notice ads on the internet that seem suspiciously tailored to your recent conversations or private emails, or if you are seeing similar types of advertisements across different websites, the appropriate action would be to check your security and privacy settings and permissions on your applications and devices.

INCLUSION

JUST LIKE US: Encouraging a team to reflect on their workplace culture by mapping out all the unwritten rules and norms, then identifying which might be exclusionary without it being previously explained to be exclusionary.

Pointing out that whenever we observe and find ourselves in a homogenous group (be it racial, education level, age, religion, political leaning or any other situation where you are in a group where there is little or no diversity), a proactive response is to seek diverse perspectives to challenge and broaden the groups collective vantage point and understanding.

A.I.

STATEMENTS

Shoshana Rosenberg

PRIVACY

Training the AI by feeding it various customer interaction transcripts and it learns to identify which portions of text or interaction typically contain or yield sensitive personal data by noticing common patterns and clusters.

INCLUSION

An AI model is given employee feedback text and identifies themes, learning to recognize sentiments and language indicative of exclusive or inclusive workplace environments.

3

SEMI-SUPERVISED LEARNING

"Do not turn the page until you have identified which of the following categories each example belongs to."
This learning uses a combination of a small amount of labeled data and a large amount of unlabeled data to improve learning accuracy.

A combination of direct instruction and self-guided exploration.

PRIVACY

JUST LIKE US: Providing a basic primer on common data privacy terms and then having individuals evaluate a series of less clear-cut cases to determine if they could potentially be privacy risks.

Discussing with friends and family that if they notice that they receive unexpected emails from a an unknown or even a seemingly known sender with some accurate information and a request for them to take any action, it could indicate a phishing attempt and it is worth verifying the source through another medium.

INCLUSION

JUST LIKE US: Starting with clear examples of inclusive language and actions, then asking individuals to assess new scenarios where the inclusive or exclusive nature is not as clear.

Where an organization reviewing its own DEI mandates notices underrepresentation in leadership roles and this prompts a mentorship program to support the career advancement of minority groups, proactive measures are being taken based on limited but important information.

A.I. STATEMENTS

Shoshana Rosenberg

PRIVACY

The AI that starts with a baseline definition of and ability to identify basic personal data and is then exposed to a larger dataset where it begins to recognize less obvious personal information that would meet that definition, like unique device IDs or inferred information.

INCLUSION

The AI is provided with a limited number of examples where non-inclusive language has been identified and learns to extrapolate those findings to identify similar language across dozens of communications and policies without labels.

4

REINFORCEMENT LEARNING

"Do not turn the page until you've completed the simulation activities successfully." The AI learns to make decisions by performing actions and receiving rewards or penalties. (Positive reinforcement would then be applied for not turning the page. “Congratulations! You may proceed.”)

Learning by trial and error through rewards and penalties.

PRIVACY

JUST LIKE US: Using a game where participants get points for correctly identifying potential privacy risks in everyday online interactions, reinforcing the learning through immediate feedback.

Teaching that if attempts to access your accounts from unfamiliar locations are flagged, then reinforcing your privacy and digital security by enabling stricter verification processes for unrecognized devices is recommended.

INCLUSION

JUST LIKE US: A simulation where positive feedback is given when participants choose actions that foster inclusivity, helping them learn appropriate responses in real-time.

Advising that if inclusive efforts are met with resistance, consistently presenting the benefits of diversity and correcting misconceptions can gradually change attitudes, as can supporting and encouraging inclusive behaviors where you see them occur.

A.I. STATEMENTS

Shoshana Rosenberg

PRIVACY

An AI can be trained and programmed by receiving rewards when it correctly identifies and secures personal data leaks in a simulated environment, with negative consequences when it fails to do so.

INCLUSION

An AI receives feedback as it navigates through a simulation involving various workplace and community scenarios, learning from positive reinforcement when it makes choices that support diversity, equity and inclusion.

TRANSFER LEARNING

5

"Do not turn the page until you understand how the previous lessons can be applied here."Transfer learning involves taking knowledge from one domain and applying it to another.

Applying knowledge from one area to another related area.

(SO META,RIGHT?)

PRIVACY

JUST LIKE US: Hosting workshops that leverage individuals' understanding of personal privacy in the physical world to navigate privacy in digital spaces, transferring concepts from one domain to another.

When learning about data breaches or phishing attempts in the news or through acquaintances, applying the leaned need for vigilance to your personal information management by updating software or settings can prevent similar vulnerabilities.

INCLUSION

JUST LIKE US: If a group of employees have been effectively trained to interact with and accommodate people with visible disabilities in the workplace, the same underlying principles can be used to help that team create an inclusive environment for people with invisible disabilities.

Noting to your team or leadership that where an inclusive policy from one department or region effectively improves team cohesion, replicating that policy in another department or region can be beneficial.

STATEMENTS

Shoshana Rosenberg

A.I.

PRIVACY

An AI retrained on recognizing personal data in English is given a small dataset in another language, adopting its privacy protection abilities/utility to a new context.

INCLUSION

An AI model that has been successfully trained to detect exclusionary language in online forums is adapted to monitor and suggest improvements for corporate communications.

ACTIVE LEARNING

6

"Do not turn the page until you decide which queries will best improve your own performance. "The AI actively queries the user to label new data points that would be most beneficial for learning.

Choosing specific challenges or tasks that will most enhance one's learning.

PRIVACY

JUST LIKE US: Setting up a forum where individuals are encouraged to ask questions about specific privacy scenarios they encounter, facilitating a more personalized understanding of privacy risks.

If you hear about a new form of phishing or online scam, actively seeking out and sharing information about it with your friends and professional network can prevent or minimize its reach and effectiveness.

INCLUSION

JUST LIKE US: Role-playing exercises where participants are active agents in creating inclusive scenarios, practicing responses to situations as they arise.

If you notice subtle exclusion in team communication, actively asking for and supporting input from quieter or seemingly disregarded team members can promote a more inclusive dialogue.

A.I.

STATEMENTS

Shoshana Rosenberg

PRIVACY

To teach an AI model about privacy, one would present datsets with examples of private and non-private information. The model would make predictions instead of labeling, with a human providing feedback on the model's predictions, with the AI actively asking the human where it found ambiguous cases to best facilitate its learning.

INCLUSION

An AI could engage with a social interaction simulation with diverse avatars representing different lived experiences. The simulation could provide feedback based on pre set diversity and inclusion metrics, allowing the AI to learn from the outcomes of its actions.

7

ENSEMBLE LEARNING

"Do not turn the page until you've reviewed the predictions from multiple models." Several learning algorithms are used together to obtain better predictive performance, and ideally the AI (and we) can understand how each model contributes to the final answer or decision.

Drawing from multiple sources or perspectives to reach a conclusion.

PRIVACY

JUST LIKE US: Hosting a discussion group that uses multiple case studies from different sources to piece together a comprehensive understanding of modern privacy challenges.

Combining different potential signs of identity theft from different aspects of your life, like unexplained bank or credit card transactions and missing mail should prompt a comprehensive identity theft check.

INCLUSION

JUST LIKE US: Facilitating a series of team-building activities designed by different facilitators to address various aspects of DEI, creating a holistic learning experience.

Observing successful grassroots affinity groups based on common experiences or identities in various departments (or other companies) can lead to the implementation of an integrated, company-wide, affinity group employee resource group program.

A.I. STATEMENTS

Shoshana Rosenberg

PRIVACY

Several AI models each trained on different types of simulated personal data (financial, health, online behavior) could be used to work together to provide a comprehensive assessment of privacy risks in large datasets.

INCLUSION

DIfferent AI systems trained on recognizing verbal, visual and visual cues that indicate inclusive environments could be used to combine insights to evaluate the overall inclusivity of multimedia content.

FEW-SHOT LEARNING

8

"Do not turn the page until you've examined these few examples of rare bird species." Here, an AI is trained to learn from a very limited amount of data, emphasizing generalization from few examples. *"Having studied the rare examples, turn the page to attempt identifying the species in different contexts with minimal guidance."*

Quickly grasping new concepts with minimal exposure or examples.

PRIVACY

JUST LIKE US: A challenge where participants are shown a small set of examples where privacy was compromised and then asked to extrapolate and identify potential new threats.

Alerting individuals that if a few cases of identity theft have occurred in their community, they should be vigilant and consider measures like credit monitoring services.

INCLUSION

JUST LIKE US: Interactive sessions where a few key principles of DEI are explained, and participants are then invited to apply these to a range of different, and perhaps less obvious, scenarios.

If a couple of team members identify that they feel excluded by certain practices or circumstances, addressing their concerns promptly can proactively prevent a wider negative impact on team morale and culture for others who may feel the same way.

A.I.

STATEMENTS

Shoshana Rosenberg

PRIVACY

An AI could be presented with a small number of carefully selected examples where data privacy has been compromised and learns to generalize these principles to detect privacy issues in larger datasets or proposed scenarios.

INCLUSION

A model could be given a few exemplary cases of inclusion best practices and successful DEI programs within leading companies and learn to extrapolate and identify similar practices or gaps amongst practices in other organizations.

META LEARNING

9

"Do not turn the page until you've completed the learning strategy exercises." Sometimes known as "learning to learn," an AI trained this way can apply knowledge learned from previous tasks to new and **related** tasks.

Reflecting on and improving one's own learning process.

(EVEN MORE META, I grant you.)

PRIVACY

JUST LIKE US: Teaching individuals how to learn about privacy by equipping them with foundational knowledge and the skills and good resources to help them understand new data protection laws and technology issues as they emerge.

Upon realizing the value of privacy in one context, like social media settings, we can extend that understanding to other areas, such as smart devices and newer cars, and their data collection practices.

INCLUSION

JUST LIKE US: Workshops for teams that focus on how to learn more about DEI and about creating inclusive spaces, empowering individuals to stay informed about evolving DEI best practices.

After acknowledging the success that an inclusion initiative had in changing feedback, understanding that adapting that understanding to improve and refine policies company-wide can be beneficial.

STATEMENTS

Shoshana Rosenberg

A.I.

PRIVACY

An AI system trained via Meta learning could be exposed to a series of different tasks related to different aspects of privacy law and data protection, allowing it to learn a general strategy for new privacy-related tasks it encounters.

INCLUSION

An AI could be exposed to various DEI training tasks, such as detecting biases or promoting equality to train it to develop a strategy to apply these learnings to new DEI challenges across industries.

10

SELF-SUPERVISED LEARNING

"Do not turn the page until the missing words in the text are filled in correctly."With this type of learning/training, AI learns to understand data by predicting any missing parts of the input data.

Setting one's own tasks and evaluating one's own progress.

PRIVACY

JUST LIKE US: Leading an activity where participants analyze a series of online user experiences to identify potential privacy issues without being told what to look for, promoting an intrinsic understanding.

If for any reason you start to question the security of your information when using certain services, it is wise to conduct personal research into alternative services that prioritize user privacy.

INCLUSION

JUST LIKE US: Setting up group discussions where people draw on their personal experiences to identify subtleties in language and behavior that might contribute to a non-inclusive environment.

Where you sense or discern a consistent lack of diverse representation in the media you consume, self-curating a more varied selection can broaden your perspective.

A.I.

STATEMENTS

Shoshana Rosenberg

PRIVACY

An AI model pretrains itself to predict the next word in a sentence and then uses this ability to identify when a name, phone number, or other private information should not be predictively generated in text.

INCLUSION

The AI trains itself on historical data regarding company or team performance and feedback, learning to infer what inclusive team dynamics look like by predicting future team success.

FEDERATED LEARNING

11

"Do not turn the page because you stay on your page and others get their own page." This method has the model learning from decentralized data sources without directly sharing or combining that data - **preserving Privacy**.

Learning from the collective wisdom of a group while keeping individual contributions private.

PRIVACY

JUST LIKE US: Different local community groups could each hold privacy workshops tailored to their specific needs and concerns, with the insights and feedback aggregated and shared as collective intelligence and used to develope a comprehensive privacy education program that benefits all groups involved.

When different apps on your phone have varying privacy policies and controls, establishing a common set of personal privacy standards that you apply across all apps can provide a uniform level of security.

INCLUSION

JUST LIKE US: If different organizations participate in a collaborative forum or Chatham house rules roundtable where they share generalized outcomes and best practices from their respective diversity initiatives, each organization can maintain the privacy of its own information but the collective wisdom can be used to update a shared set of guidelines for all organizations to work from or implement.

Taking on privacy-centric inclusion feedback from individuals through a trusted third party where their privacy rights are maintained, where the organization itself receives a heat map of issues that they have to address holistically.

A.I.

STATEMENTS

Shoshana Rosenberg

PRIVACY

AI models could be trained on user devices to recognize personal data without the data ever leaving the device, ensuring privacy while still benefitting from shared learning about overarching trends and insights.

INCLUSION

AI could assist a large multinational organization by identifying effective inclusion practices across its global footprint based on local datasets being used to train isolated local AI models based on local cultural and organizational variables. Instead of pooling the data centrally, each office sends only their models parameters to a central server that aggregates the parameters to create a global model.

12

MULTI-INSTANCE LEARNING

"Do not turn the page without understanding the context of grouped examples." *(Yes, that one is a reach, for sure.)* The key here is:

Understanding the general characteristics of a category without studying individual examples in detail.

PRIVACY

JUST LIKE US: Imagine a workshop where participants are given different case studies of data breach incidents. Each one represents one instance and contains varying details about the breach circumstances. Participants must analyze these case studies to determine common patterns and subtleties that can indicate a breach.

When certain online forms or applications ask for personal information in excess of what is relevant or necessary, being able to discern which are necessary and which are intrusive can help maintain your privacy.

INCLUSION

JUST LIKE US: Setting up a series of team building exercises where each team encounters various scenarios involving inclusion challenges. Teams must work through these challenges to find common underlying principles of inclusive behavior and commonalities between exclusive behaviors.

Reading multiple case studies from different companies that implemented inclusion feedback mechanisms or policies could allow one to compare outcomes and strategies across the instances to identify common effective practices and pitfalls, even if some information is incomplete or some case studies are outliers.

STATEMENTS

Shoshana Rosenberg

A.I.

PRIVACY

A cybersecurity AI could be tasked with detecting phishing attempts across different email clients within an organization. Each email is considered a "bag" of features such as sender info, embedded links, and language patterns. Only some emails are labeled as safe or phishing and the AI learns to classify based on the labels AND patterns across the features.

INCLUSION

An AI system could be developed to evaluate workplace culture by analyzing employee feedback across multiple departments - with the feedback instances being grouped by topic. The AI learns to recognize inclusive or exclusive culture indicators across different topics, helping the organization target group wide changes where they are needed most.

13

LEARNING WITH HUMANS IN THE LOOP

"Do not turn the page and ask questions if you are unsure of the answer or correct action and need further guidance." This learning model is designed to solicit and incorporate human expertise when its confidence in a prediction is low.

Enhancing one's skills through iterative feeback and expert intervention when faced with uncertainty.

PRIVACY

JUST LIKE US: If a tech firm set up a user council with broad representation that reviews new products for potential privacy concerns, the users could identify subtle privacy issues that might arise from various user interactions with the technology.

If you are part of a team or project team responsible for handling personal data on behalf of a client, when in doubt about a current or new data usage, consult with your Privacy or legal team before proceeding.

INCLUSION

JUST LIKE US: In a university seeking to ensure a rounded program, students and faculty could participate in curriculum development committees to ensure that courses include a wide range of perspectives. Input from teachers and students could lead to the inclusion of often overlooked historical figures or contributions from diverse cultures in the university's courses.

When a team member points out that a practice may be unintentionally exclusive, involving those affected to co-create a more inclusive approach is key.

STATEMENTS

A.I.

PRIVACY

Human experts can monitor an AI's privacy determinations or decisions in real time, providing immediate feedback on accuracy and continuously improving its understanding of privacy in complex scenarios.

Shoshana Rosenberg

INCLUSION

As an AI proposes language for corporate communications, DEI experts can review and adjust the outputs, training the AI on nuances of language that promote a culture of inclusion.

A.I., **Privacy** and **DEI** have a high level of interdependence and interconnectedness and are continuously evolving.

All three are tied directly to ethics, fundamental human rights, the future of work, and decision making and bias, which means:

YOU AREN'T ON THE SIDELINES OF THESE THINGS.

YOU ARE CRUCIAL TO THEM BEING WHAT THEY SHOULD.

Do you want to know more?

STATEMENTS

Shoshana Rosenberg

THE WAY I SEE IT

DIVERSITY, EQUITY, AND AI ARE THE FUTURE

AI will accelerate the interconnectivity of the world and will be deeply ingrained in all kinds of decision making processes.

DEI is essential for sustainable progress, innovation, and harmony.

Embracing DEI in the AI era is the only viable path to ensure that AI does not exacerbate or perpetuate existing biases and inequity.

INCLUSION IS THE KEY TO DIVERSITY AND EQUITY

Diversity will not be sustainable and equity not possible unless a full spectrum of the community is represented, integrated, accepted, respected and valued.

PRIVACY IS THE KEY TO INCLUSION

Inclusion cannot be fostered effectively without candid feedback and diversity identity data, both of which put indiviuals at RISK without Privacy and true anonymity, and which are too often collected without preserving their privacy rights.

DATA IS THE KEY TO AI

To ensure ethical AI, the data it is trained on must be diverse, representative, and gathered properly (with authorization and consent) and used responsibly.

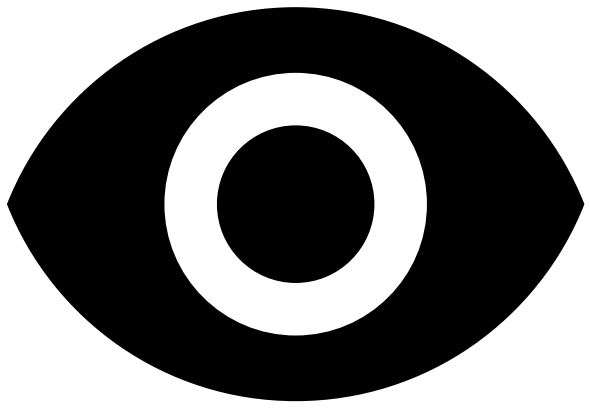
AI CAN BE AN UNPARALLELED KEY TO EQUITY

AI that is properly and thoughtfully designed with DEI principles can identify and help flag and rectify systemic disparities across any number of sectors and disciplines and processes, as well as helping to identifying gaps in the policies or tools that support them.

STATEMENTS

Shoshana Rosenberg

TRANSPARENCY, ACCOUNTABILITY, FAIRNESS, AND TRUST ARE KEY TO ALL THREE



A.I. DETECTS STATE PRIVACY

Shoshana Rosenberg