A HEALTHY NARRATIVE
FOR ARTIFICIAL INTELLIGENCE

Measuring Success for
Mozilla's Creative Media Awards Program

HARMONY LABS
311 W 43 St, 12th floor
New York, NY 10036

+1 212 966 7606
info@harmonylabs.org
harmonylabs.org

TOWARD A HEALTHY NARRATIVE
FOR ARTIFICIAL INTELLIGENCE
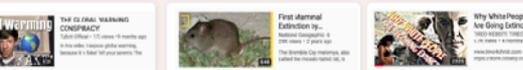
The Mozilla Foundation is working to change the dominant narrative around artificial intelligence (AI).

Success in narrative change is more than shifting support for a single policy or idea. It's about embedding and/or reinforcing a deep, shared story about what AI is, how it works, and how it should be managed.

We can't measure success without knowing what that story is, nor can we find opportunities for telling that story effectively without understanding the narrative landscape.

WHAT IS
NARRATIVE?
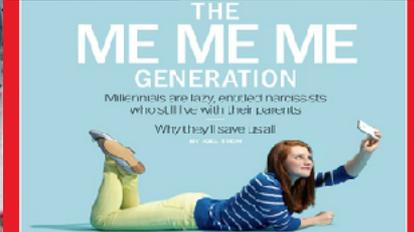
Narratives are story patterns or archetypes: the characters, emotions, values, and/or endings that occur in consistent patterns across stories about one social issue.

We find stories in all the domains of culture and the data they generate: news, social media, television, film, music, etc. We use survey to understand the narratives in people's minds.

Stories about different topics in AI, like criminal justice or business, could share a single narrative, such as "AI is a tool of tyranny we ought to fear."

HOW DO WE
IDENTIFY NARRATIVES?

To understand the narrative landscape around AI and to identify a healthy target narrative, we started by analyzing Mozilla survey data and news content data.

We built a natural language processing pipeline to read thousands of survey responses and news stories to find the story patterns that naturally occur together, which we call narratives. Then human annotators looked at the clusters, read many of the stories, and classified them into groups.

1
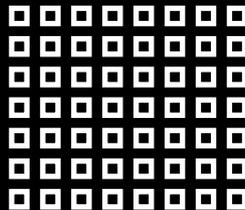
Open-ended responses to a Mozilla survey question: "Which two (2) words best describe how you feel about AI in the world right now? Please tell us why."

2
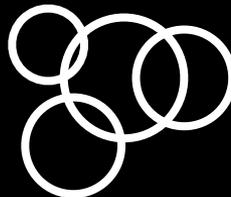
Computers read responses to learn which ideas, emotions, values, and metaphors occur together. Then suggests some naturally occurring response clusters.

3

Humans review the stories in each cluster and verify and assign meaning to the narratives.

AI NARRATIVES IN THE SURVEY

Natural language processing surfaced 5 groups of stories containing consistent narrative elements. Each narrative had something to say about:

- Who or what AI is
- What emotion the audience feels about AI
- What happens in the story if it goes right—the happy ending
- What happens in the story if it goes wrong—the unhappy ending

SURVEY

AUGMENTED INTELLIGENCE

CAUTIOUS & CURIOUS

TOOL OF TYRANNY

EVIL INTENTIONS

ROBOT OVERLORDS

NEWS
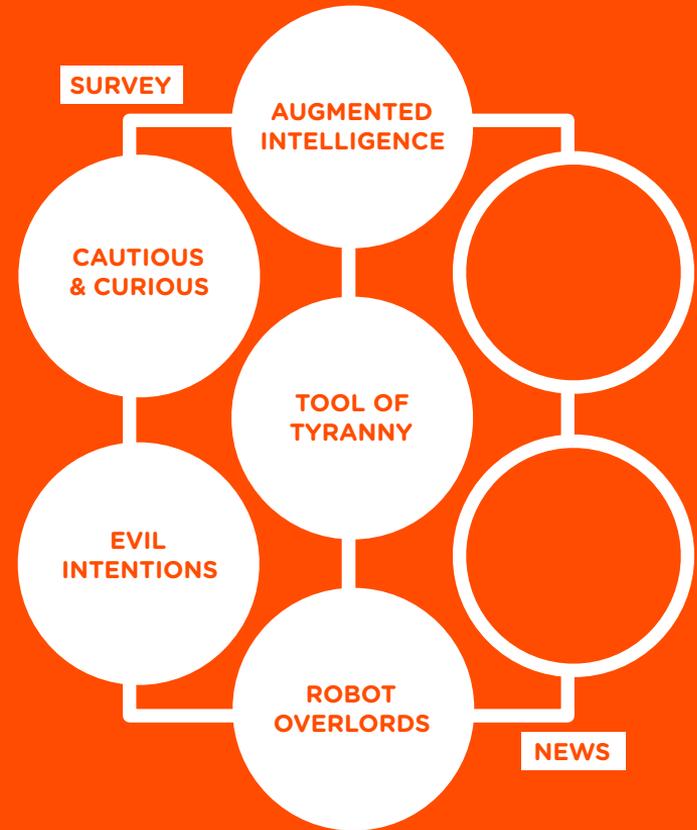
NARRATIVE 1
AUGMENTED INTELLIGENCE

AI is simply a tool—a computer program created by humans. We are calm and optimistic about AI, but aware of the risks.

If we are successful, AI will augment our ingenuity and liberate us from repetitive tasks. It will take vigilance and effort to overcome human error, algorithmic biases, and limitations in our parameters and data.

NARRATIVE 1
AUGMENTED INTELLIGENCE

"I think Artificial is not the term that we are striving for, instead it's Augmented. Augmented Intelligence . . . using our own and augmenting it."

"It can remove tedious and error-prone tasks."

"Programer bias creeps into the code, hard wiring biases into systems."

"Computers become more useful as hardware and software evolve."

NARRATIVE 2
TOOL OF TYRANNY

AI is a spy used by governments and corporations to mislead us, invade our privacy, and limit our freedoms. We are afraid of AI and angry at the governments and businesses that use it against us.

If we get this right, we might be able to defend ourselves. But the die are already cast, and we're probably going to suffer lost jobs, lost freedoms, and a degraded way of life.

TOOL OF TYRANNY

"More ways that businesses, governments, hackers, will have to access more of my personal information."

"Corrupt governments and corrupt capitalists will use this against an individual's right to privacy."

"There goes all our privacy with face scanning, just like the Chinese government."

"The loss of privacy manipulation and jobs."

ROBOT OVERLORDS
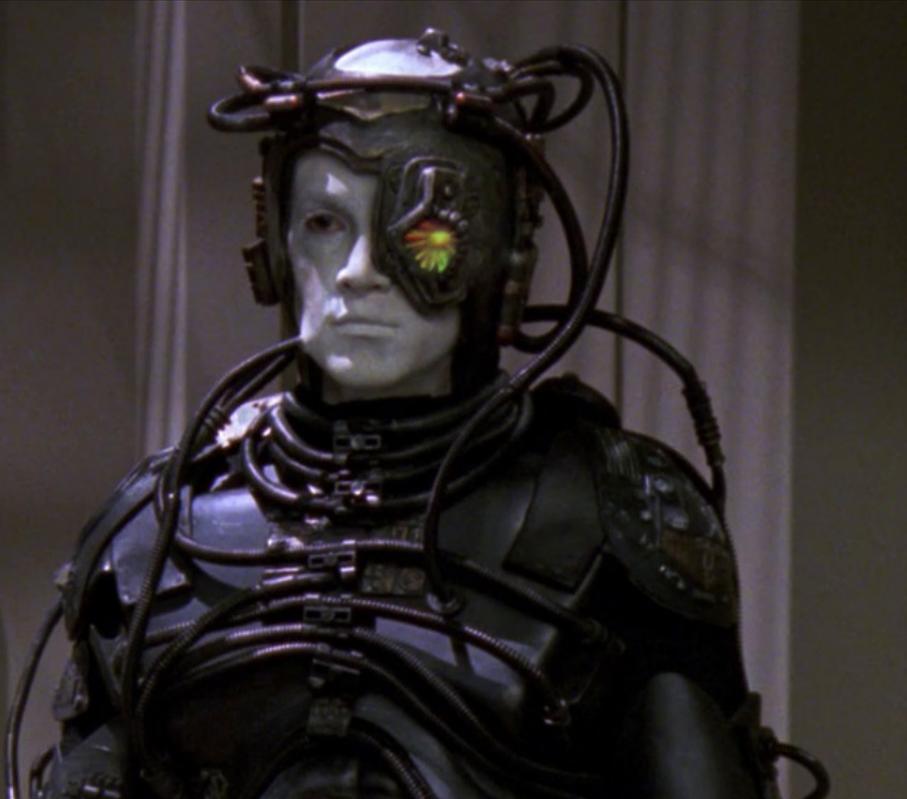
AI is created by people too lazy to do simple tasks, but it is an agent with its own agenda, capable of thinking for itself. Because we have all seen *Terminator*, we are afraid of AI.

Our best hope for the future is that people who can still think for themselves will resist the invasion of AI. If this goes wrong, the robots will make all our decisions for us.

NARRATIVE 3
ROBOT OVERLORDS

"My concern summed up in one word . . . TERMINATOR."

"AI is something that is coming for our children and grandchildren!"

"People need to remember how to exercise and do some things themselves."

"I don't like the idea of leaving a machine decide."

NARRATIVES 4 + 5
UNDEVELOPED

CAUTIOUS & CURIOUS and EVIL INTENTIONS were two perspectives that came up frequently in the survey, but in our view do not constitute narratives. Instead, they are stories survey respondents tell when they haven't really thought much about AI, relying instead on general beliefs about how the world works.

These beliefs are important because they are held by groups whose perspectives might be relatively easily shifted with supplemental information about AI; although they are not necessarily the groups most likely to report low "awareness."

AUGMENTED INTELLIGENCE

CAUTIOUS & CURIOUS

TOOL OF TYRANNY

EVIL INTENTIONS

ROBOT OVERLORDS
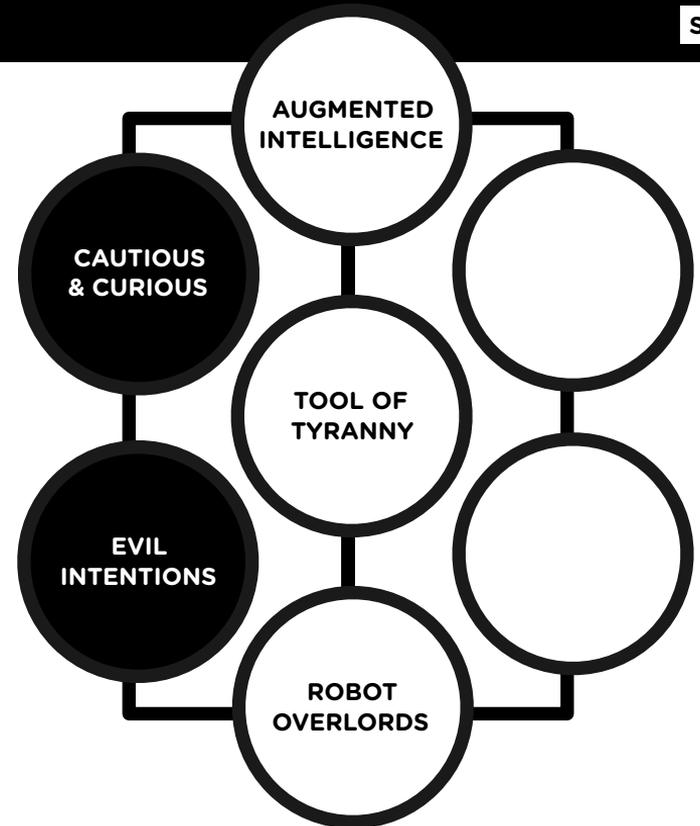
CAUTIOUS & CURIOUS

AI is an unknown technology, and the future seems filled with promise. We are excited about AI and just a little nervous. Innovation may lead to tremendous progress, but we should probably be aware that there could be "unknown complications."

"I am hopeful that it will benefit us all and curious to see how it will be implemented."

"With all new things, some good and some bad come with it. So let's hope for the good."

EVIL INTENTIONS

AI seems like a powerful tool that can be used to help or harm. On balance, however, we are afraid of AI, because the world is full of bad people seeking power, and those people will misuse the tool. If we get this right, we will be safe. If we get it wrong, the result could be devastation.

"AI can be used in beneficial ways to mankind in general, BUT IT WILL BE ABUSED IN EVERY WAY POSSIBLE!"

"Bad people can use it with devastating effects"

AI NARRATIVES IN THE NEWS

These narratives are interesting, but are they "real"? Do they exist in the world? Are they prevalent in the culture? They might just be outgrowths of a survey format, a statistical analysis, or a respondent sample.

To find out, we ran a completely separate analysis of the narratives in more than 7,000 news articles that mentioned "artificial intelligence" or "machine learning."

3 of the narratives people shared on the Mozilla survey—Augmented Intelligence, Tool of Tyranny, and Robot Overlords—also appeared in the news articles.

SURVEY

AUGMENTED INTELLIGENCE

CAUTIOUS & CURIOUS

TOOL OF TYRANNY

EVIL INTENTIONS

ROBOT OVERLORDS

NEWS
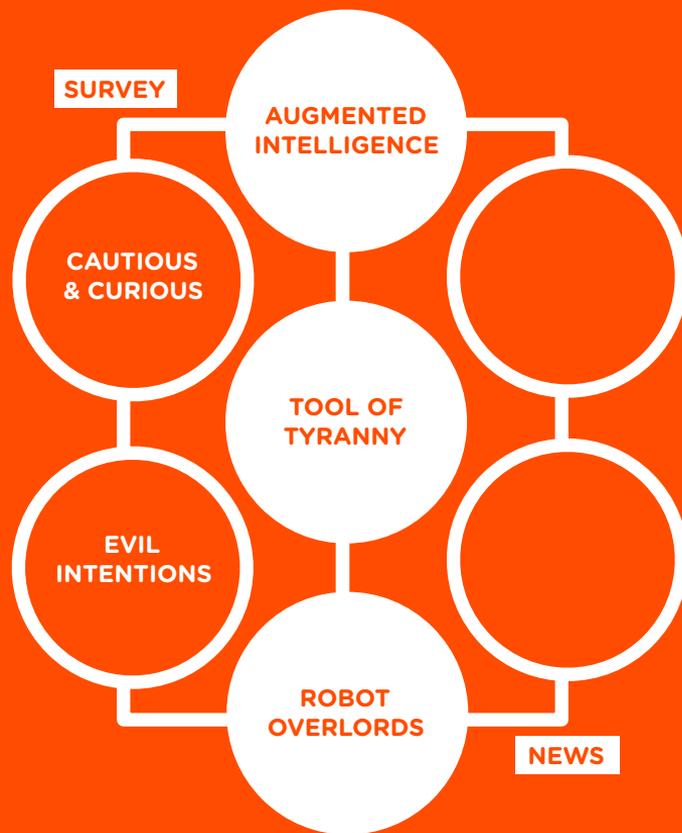
NARRATIVE 1
AUGMENTED INTELLIGENCE

"New technologies . . . are either demonized and immediately banned as a vital threat to civil liberties or warmly embraced as a new way to improve safety and convenience. Decisions are based far too much on narrative and perception, and far too little on facts and reason." —Fast Company (1/13/20)

"An artificial intelligence (AI)-based Google model has left radiologists behind in spotting breast cancer" —Business Insider (1/2/20)

" . . . A global model for clean energy and green economy by using the disruptive technologies of the fourth industrial revolution such as: artificial intelligence" —Emirates247 (1/5/20)

When appearing in the news, this narrative is often associated with descriptions of European projects.

Optimism in this coverage is associated with AI's ability to take on complex tasks in healthcare, self-driving cars, and green energy. Human rights also make an appearance, emphasizing that AI is by people and should be for people.

NARRATIVE 2
TOOL OF TYRANNY

"Western allies must coordinate against the tyrannical uses of artificial intelligence pioneered by the Chinese communist party" —Washington Examiner (1/23/20)

"Element AI is leading a human rights impact assessment of the proposed Alphabet Inc's Sidewalk Labs smart city development" — Financial Post (1/16/20)

". . . Serious talk about the millions of jobs lost to automation and artificial intelligence and the dark outlook for American jobs." —Baltimore Sun (2/11/20)

"Deepfakes is the most terrifying development in artificial intelligence that I've ever seen." —The Money Show (1/22/20)

This narrative appears in the news as well, where the threat to the U.S. from China in particular is an important theme.

In the news, as in individual open-ended survey responses, job loss to automation is connected with this narrative.

NARRATIVE 3
ROBOT OVERLORDS

"In 50 years we will look back, and it'll represent a time where humans were being invaded by artificial intelligence. Not a bad thing. Just what it is." —NJ.com (1/18/20)

"Soon, Spotify was giving me playlists via machine learning quicker than any crush could." —The Stanford Daily (2/12/20)

". . . dead-hand system with decision-making by artificial intelligence, after human decision-makers are killed has been proposed." —ForeignPolicy.com (2/11/20)

"The problem with cyborgs is they are much harder to catch and detect". —Associated Press (2/7/20)

This narrative appears in the news as well where it is characterized by AI's agency and personification and by language like "cyborgs" and "droids."

Some of this news coverage is actually about science fiction movies and books. Where real technologies appear, deepfakes figure prominently.

ADDITIONAL NEWS NARRATIVES

Two narratives were unique to the news: MARKETS & MONEY and SECURITY & TECHNOLOGY. These narratives feature institutions—government and business—as protagonists, as opposed to people.

The SECURITY AND TECHNOLOGY narrative, in particular, contained much of the content concerning AI rulemaking. That "rule" was more prominent in the institutional narrative than in AUGMENTED INTELLIGENCE suggests that there is still opportunity to connect individual optimism for efficiency and innovations to practical approaches to AI regulation and management.
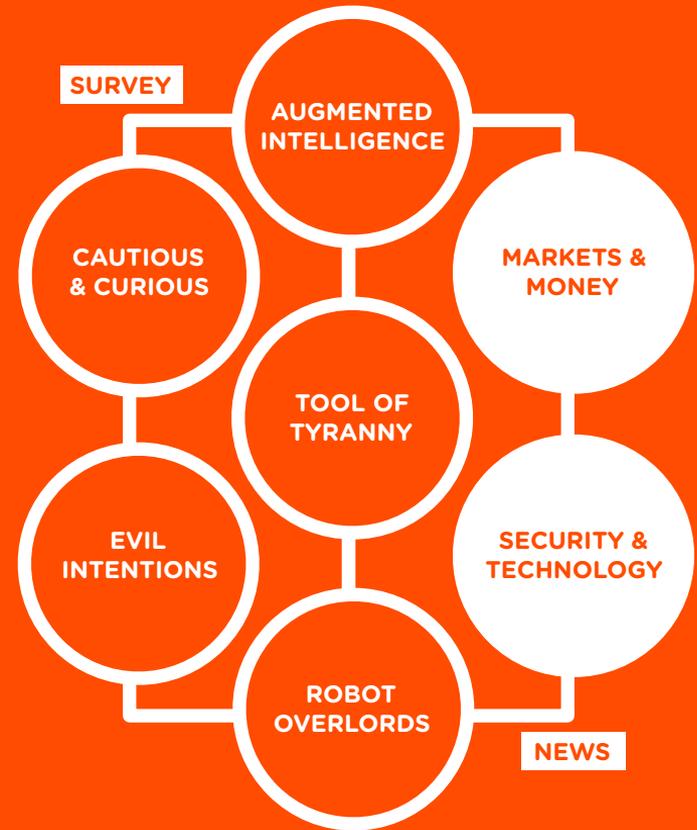
"Our proven AI, machine-learning technology and retail industrial engineering help you accelerate ROI" —Business Insider (1/28/20)

"Increasing revenues . . . in financial services could be greatly enhanced by deploying more accurate deep learning models, but . . . this would be arduous to explain to regulators." —Business Insider (1/10/20)

"Analysts warn that western defense capabilities must adapt to modern threats." —VOA News (2/15/20)

"Britain's most senior police officer . . . called on the government to create a legal framework for police use of new technologies." —Reuters (2/24/20)

AI is a product with immense money-making potential. We feel excited. If we get this right, we will make a mint! If this goes wrong, we might be regulated out of existence.

AI is a weapon that requires stringent regulation. We feel cautious and determined. If we get this right, we will create rules that keep us safe. If we get this wrong, our enemies will outmaneuver us.

MARKETS & MONEY

SECURITY & TECHNOLOGY

NEWS

TARGET NARRATIVE
AUGMENTED INTELLIGENCE

AUGMENTED INTELLIGENCE is the naturally occurring narrative that is closest to Mozilla's healthy AI goals. Relatively informed about what AI is, people who hold this narrative are aware of both potential benefits and risks of the technology. To measure whether a specific media project reinforces this narrative and moves them away from other narratives, we can ask questions like:

- In your opinion, is AI going to make our world better, worse, both, or neither?

- How much do you know about AI?

- Which of these comes closer to your view, even if neither is exactly right: AI is technology made by humans, and humans control AI; or AI has the potential to develop a mind of its own and control humans.

- Like all technology, AI has risks and benefits. To what extent do you agree that there are ways to design AI to ensure that humans remain in control; or there are ways to regulate AI to ensure the technology benefits people and doesn't hurt them.

TESTING FOR SUCCESS
TRIAL DESIGN

With baselines established for core narratives, the next part of the study will aim to measure how effective specific media interventions are at moving people towards the target narrative. We'll use randomized control trials, to produce insights into the efficacy of three projects:

- *In Event of Moon Disaster* by Francesca Panetta and Halsey Burgund
- *Deep Reckonings* by Stephanie Lepp
- *TheirTube* by Tomo Kihara

TREATMENT GROUP

survey respondents see Mozilla projects

CONTROL GROUP

survey respondents see neutral content

BOTH GROUPS ANSWER NARRATIVE SURVEY QUESTIONS

content "works" if the treatment group agrees more with Augmented Intelligence than control group
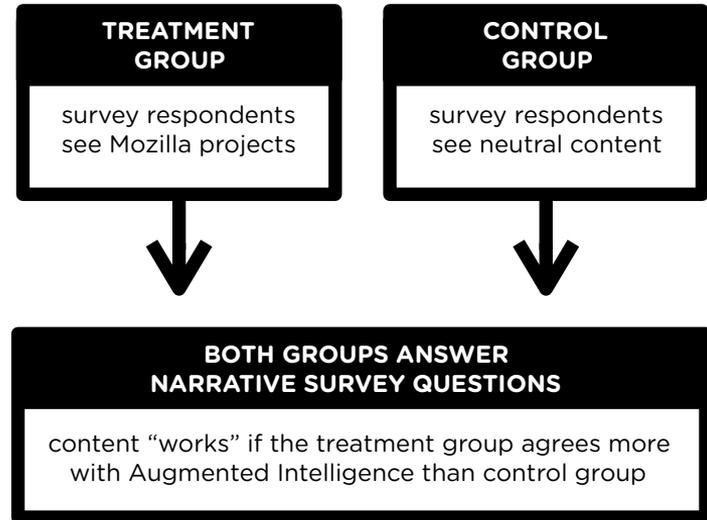
THE RESULTS

We found that these types of media projects did influence public understanding and emotions towards AI, though sometimes in ways that ran counter to Mozilla's larger theory of change that seeks to move people towards an understanding of "healthy AI."

1 The "optimistic but realistic" AUGMENTED INTELLIGENCE narrative is pretty rare

**2** CMA media didn't increase the AUGMENTED INTELLIGENCE narrative

3 But CMA media did change AI narratives generally

RESULT 1
AUGMENTED INTELLIGENCE NARRATIVE IS RARE

It was rare for participants to endorse the healthy narrative, i.e., to choose answers that align with AUGMENTED INTELLIGENCE values for every question across all conditions.

Only about 9% of participants endorsed this narrative fully. Even allowing for some ambivalence about both the positive and negative impact of AI, only 16% of participants endorsed this healthy narrative.

Certain groups endorse the healthy narrative more than others—technical jobs, college educated, or men. But belief in an AI that can improve our lives without posing unacceptable risks is rare even among technologists (14%).

Endorsement of Healthy AI Narrative Components

Humans Can Control AI

Human Control & Govt Can Regulate AI Effectively

Human Control & Govt Regulation & It Can make the World Better

Augmented Intelligence Narrative
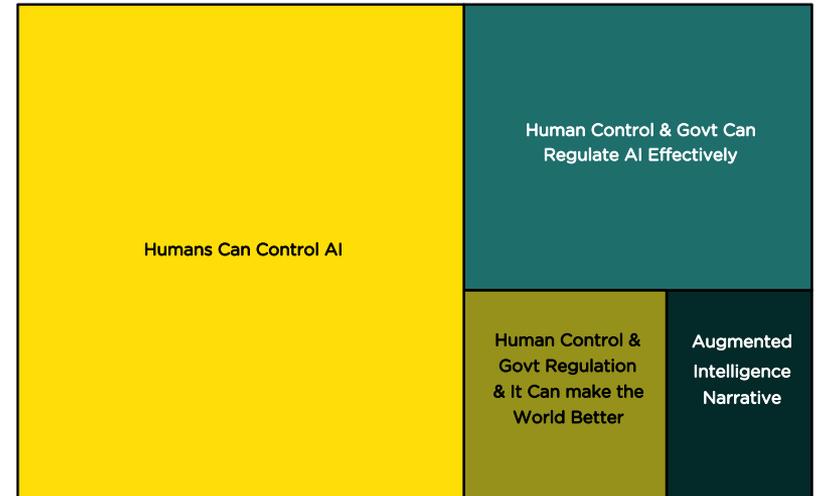
MEDIA DIDN'T INCREASE HEALTHY NARRATIVE

Each video used different tactics to articulate the impact of AI on society: *In The Event of Moon Disaster* imagines a universe in which the first astronauts never landed on the moon, *Deep Reckonings* portrays Brett Kavanaugh apologizing for his actions, and *TheirTube* envisions a tool to understand other users' YouTube recommendation algorithms.

But did any of these videos successfully move audiences toward endorsing the healthy narrative? In short, no. None of the videos significantly shifted participants closer to endorsing the AUGMENTED INTELLIGENCE narrative relative to a neutral control video.

CMA videos didn't increase belief in overall Augmented Intelligence Narrative

MEDIA DID CHANGE THE NARRATIVE

While they may not have moved participants towards the healthier narrative, all of the videos made participants feel that AI was more likely to make our world a worse place.

This effect was significant, when compared to the control, for *In the Event of Moon Disaster* and *Deep Reckonings*, which both use techniques, such as deepfakes and video editing, to make people appear to say things they've never said.


CMA videos increase the belief that AI can make the world a worse place

MEDIA DID CHANGE THE NARRATIVE

In addition, all of the videos promoted narratives that produced fear and/or imagined an unhappy ending for humanity, with increased insecurity about whether humans could maintain control over AI. Specifically, participants gravitated towards the ROBOT OVERLORDS narrative and away from the AUGMENTED INTELLIGENCE narrative.

This effect was significant for *Deep Reckonings*, which uses AI techniques to make it appear as though Brett Kavanaugh is delivering an apology speech.

CMA videos decrease the belief that humans can control AI
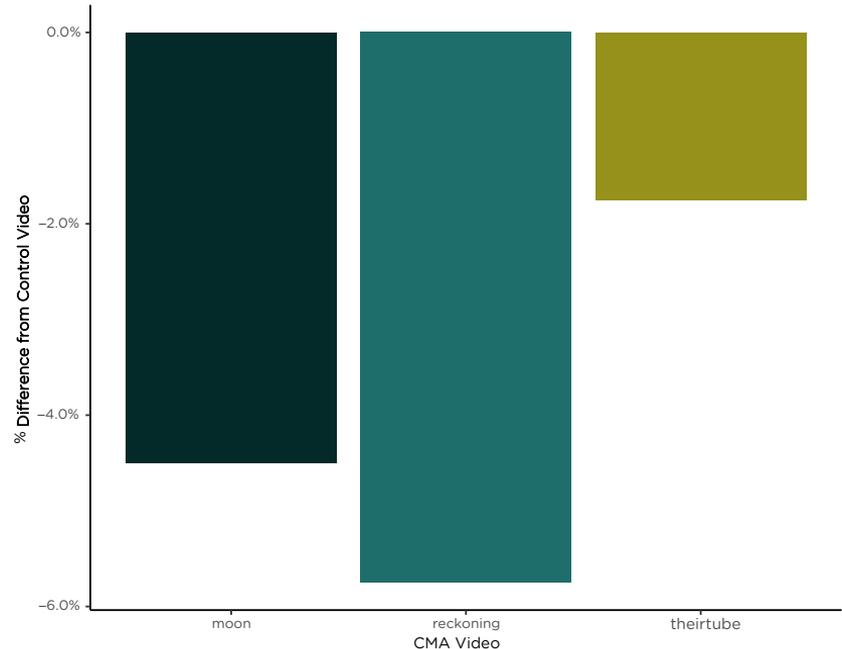
RESULT 3
MEDIA DID CHANGE THE NARRATIVE

Finally, and perhaps most importantly, all of the videos increased the belief that governments can effectively regulate AI to ensure that technology helps people, leading participants away from the TOOL OF TYRANNY narrative.

CMA videos increase the belief that the government can regulate AI effectively
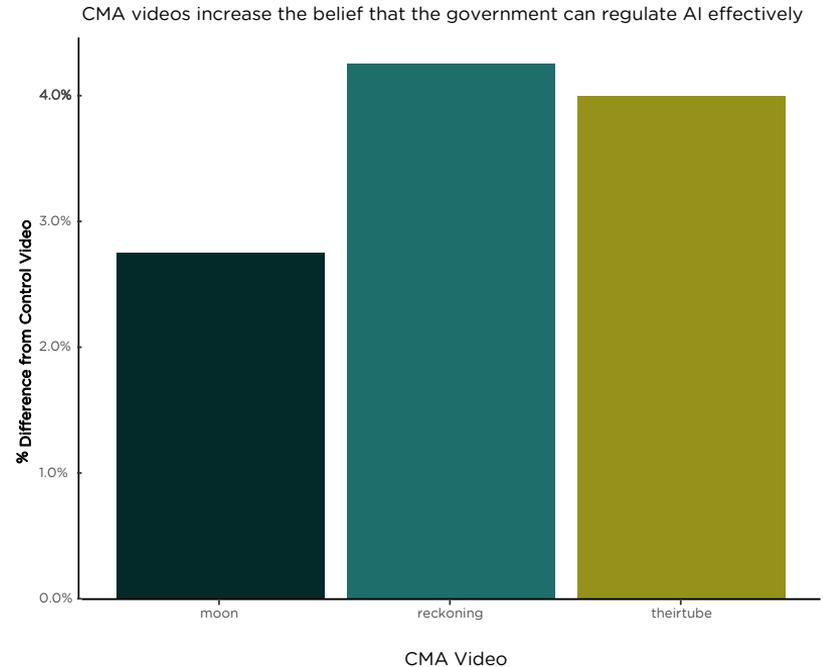
THE UPSHOT
OPPORTUNITY FOR ADVOCACY & AGENCY

While none of these videos moved people to wholeheartedly embrace the healthy narrative, they did help to rewrite it to read more like:

> AI is dangerous and could make the world a much worse place. We might not be able to control it, but we have to try. We should work to make rules to ensure that AI is not used to hurt people.

While this narrative is fear-based, it does promote regulation, which potentially leaves open the opportunity for advocacy.

The projects weren't selected with the explicit goal of moving people towards a healthy AI narrative, so this is not altogether surprising. This can be addressed in future cohorts by selecting for a narrative approach that promotes both agency and optimism.

SO WHAT ABOUT COVID?

The truth is that we don't know yet. One of the groups of news articles the clustering detected held AI and COVID content.

However, that cluster also contained some unrelated content from China and other "garbage" that suggests that the narrative is not fully formed.

The narrative opportunities and risks are emerging, though.

**OPPORTUNITY
TO SHOW HOW AI SUPPLEMENTS HUMAN JUDGMENT**

"Amazon now says that it is detecting price gouging by comparing prices submitted by sellers with historic prices offered in the marketplace. It is also using machine learning to aid with the human monitoring."
—CNBC (03/06/20)

**RISK
OF CELEBRATING SURVEILLANCE**

"A new website hopes to help, by shouting at you if you touch your face." —News Break (03/10/20)

**INCLUDING
GOVERNMENT OVERREACH**

"In Beijing, China, subway passengers are being screened en masse for symptoms of coronavirus by artificial technology and temperature scanners."
—Insider (02/11/20)

REMEMBER: PEOPLE ARE CONFUSED
ABOUT WHAT AI IS

A 2019 survey showed that most people thought technologies with agency, like robots and self-driving cars, are "artificial intelligence," but computer vision, chat bots, and recommendation models are not.

To connect all the relevant technologies, we can talk about "automation" or refer to specific applications.
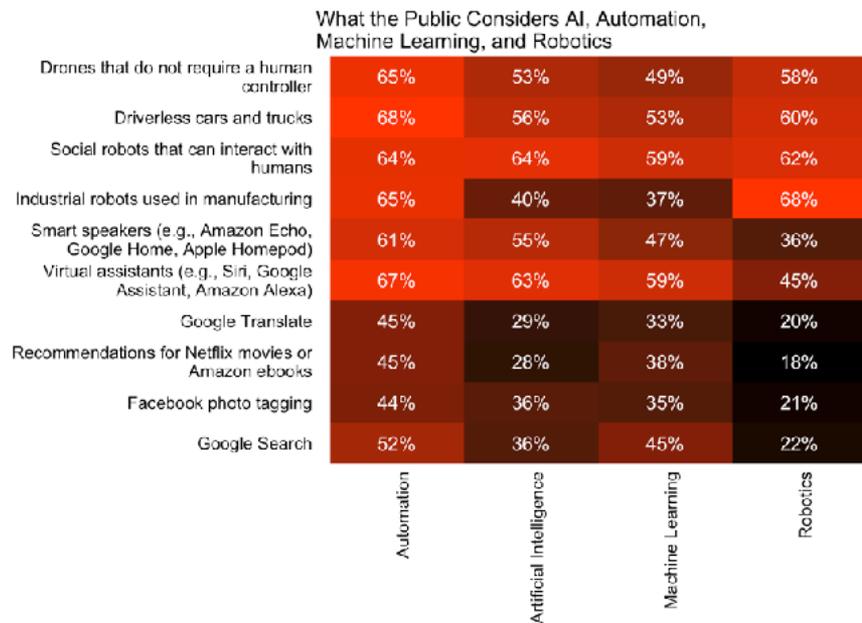
What the Public Considers AI, Automation, Machine Learning, and Robotics

	Automation	Artificial Intelligence	Machine Learning	Robotics
Drones that do not require a human controller	65%	53%	49%	58%
Driverless cars and trucks	68%	56%	53%	60%
Social robots that can interact with humans	64%	64%	59%	62%
Industrial robots used in manufacturing	65%	40%	37%	68%
Smart speakers (e.g., Amazon Echo, Google Home, Apple Homepod)	61%	55%	47%	36%
Virtual assistants (e.g., Siri, Google Assistant, Amazon Alexa)	67%	63%	59%	45%
Google Translate	45%	29%	33%	20%
Recommendations for Netflix movies or Amazon ebooks	45%	28%	38%	18%
Facebook photo tagging	44%	36%	35%	21%
Google Search	52%	36%	45%	22%

THANK YOU

HARMONY LABS
311 W 43 St, 12th floor
New York, NY 10036

+1 212 966 7606
info@harmonylabs.org
harmonylabs.org