



Framing the Social Implications of AI

Table of Contents

Introduction	3
<hr/>	
Public Thinking	5
<hr/>	
The Framing Strategy	8
Four Types of Frames	8
Summary of Recommendations	18

Introduction

With the public launch of applications such as ChatGPT, Google Gemini, and other generative AI over the past few years, AI technologies have been prominent in the public discourse. Narratives of promise, fear, ethics, and international competition around AI make this particularly complicated for everyday people to navigate. Its recent prominence in public discourse presents an opportunity to deepen the understanding of AI as a technology, its social implications, and how it can be used for the public interest.

The opportunities for AI to transform areas like health care, education, social services, and legal systems lie in how its societal purpose is envisioned. If AI systems are created to serve individual and corporate interests, existing power relations and biases in society can perpetuate harm in already marginalized communities. However, if AI systems are created in an open and transparent way that includes those communities and actively reduces harm, AI can contribute to a fairer and more sustainable society.

Despite the massive increase in public exposure to generative AI, there is still a lack of fundamental understanding about how AI works and its social implications. By examining the deeper assumptions that shape public thinking, we have identified and tested strategies for talking about the social implications of AI technology. In this strategic brief, we lay out a set of recommendations that advocates, experts, and communications professionals can use to advance this strategy:

- Define AI as a computer processing technology that, at its core, is used to identify patterns in massive amounts of data. This is critical to opening people up to a conversation about the social implications of AI.
- Distinguish AI from human intelligence and emphasize how humans are involved in AI's design, operation, and applications. This helps the public to see how people can influence AI, not just be subject to it.
- Focus on AI's reliance on data and how humans are involved in every step of AI processes. This helps people see how bias is embedded in AI systems and how it reflects the biases and power structures of our social systems.
- Talk about the harmful effects that the use of AI can have on communities, especially marginalized communities, and use concrete examples from real life. This helps people view the regulation and application of AI as issues of social justice that must be addressed through policy change.

These recommendations can be integrated with existing communications strategies to create a broader picture of the social implications of AI as a technology. The hope is that these frames, as a strategy, will enable advocates to build a collective understanding of how AI can be developed and employed for the public interest.

This brief is organized into two principal sections:

- 1. Public Thinking.** This section outlines the cultural mindsets—the implicit assumptions and taken-for-granted ways of thinking—that people use to think about AI. In this section, we highlight those mindsets, with particular attention paid to those that make it difficult to communicate the ideas that individuals working in the field want the public to understand.
- 2. The Framing Strategy.** In this section, we recommend a framing strategy that consists of four types of frames: explanations, metaphors, issues, and values. We explain what ties these frames together and offer guidance on how to apply the frames using sample messages. We also explain how this new strategy can be integrated with existing field communications.

All the frames in this brief were tested across a variety of methods, including survey experiments, focus groups, and pre/post interviews. Our [research methods supplement](#) contains a more detailed description of the methods we used to conduct the research.

Public Thinking

We have identified several key cultural mindsets that people use to think about AI technology and its social implications.

What are cultural mindsets? How do mindsets differ from public opinion?

Cultural mindsets (or mindsets, for short) are deep, assumed patterns of thinking that shape how we understand the world and how we make decisions. In shaping how we think, mindsets give rise to our beliefs, attitudes, and opinions and inform our decisions and behaviors.

In contrast to public opinion research, which tells us *what* people think about specific issues or policies, cultural mindsets research tells us *how* people think about an issue—the ways in which our tacit assumptions about the world shape how we make sense of issues, how we draw (or don't draw) connections between issues, and how we reason about needed solutions.

An important feature of cultural mindsets is that we all hold multiple, sometimes competing, mindsets. What matters is the relative strength of these mindsets and how they are applied to the issue at hand.

There are cultural mindsets that the public uses to think about AI that present important challenges for those who are communicating about the social implications of AI and about AI generally:

AI as innovation. Technology is generally understood as a tool used to make life easier and to accomplish tasks quicker, and so technology's purpose is understood as fundamentally good. AI technology is, therefore, viewed as helpful in providing a level of convenience in daily life by performing a variety of tasks, such as answering questions, providing personal assistance, and generating content. Although public attention toward generative AI has increased, people often do not understand that many of the most harmful impacts of AI come from predictive and scoring systems used in public decision-making. This mindset makes it harder for people to think critically about the negative effects of technology, including AI, when they see it as a tool to improve their lives. It also limits their ability to recognize how deeply embedded AI technology is in aspects of our lives that are not directly related to any device or app.

AI versus humans. The public sees technology broadly and AI technology specifically as standing in opposition to humankind. This division helps feed concerns that AI will replace people in jobs and that reliance on AI will make people dumber, lazier, and less able to hold human relationships. When this mindset is active, it can crowd out critical understanding of how AI amplifies human biases in problematic ways. The human-AI opposition also makes it

difficult to see the role of humans in the development, application, and interpretation of AI. This leads to a view of AI as much more independent from human intervention than it really is. Many of the most important social implications of AI can only be understood when people can clearly see the interactive relationship between the technology and humankind.

AI as a consumer product. When people think about the social implications of AI, they focus on what money can buy, which is grounded in the broader consumerist mindset that money buys quality in all areas of life. Quite simply, they view AI as a luxury product that some can, and others can't, afford. Even as AI becomes embedded in public systems and institutions, people continue to understand it primarily through a consumer lens. A consumerist perspective on AI makes it harder for people to recognize that AI brings harm to certain populations because it reflects, reproduces, and exacerbates existing systemic biases in society.

Bad actors, not rigged systems. The public perceives AI technology, and technology more broadly, as an "objective" product that does not require interpretation and is considered more accurate and less biased than humans. Because the public sees AI as an objective tool that is meant to improve people's lives, it reasons that any negative impact from the use of AI is a deviation from the norm. They blame the person using it, not the technology itself or the companies or the market that created it. Concerns about AI usage, then, become focused on controlling individual "bad actors" and thus tend to underestimate the need for government regulation of the industry and for more systemic solutions.

Collectively, these mindsets show us the cognitive landscape that the framing strategy needs to navigate. We were looking for frames that accomplish the following:

- 1. Clarify what "AI" involves and what is meant when people describe a technology as artificial intelligence.** The term "artificial intelligence" is often used as shorthand for any new, exciting, or cutting-edge technology, regardless of whether that technology is designed to mimic human intelligence. This can lead people to incorrectly conclude that some technology is AI based simply on its novelty while overlooking more commonplace uses and impacts of predictive algorithms. The challenge here is to reframe how the public understands what AI is in order to foster a fundamental change in the conversation around what AI development and use represents in society.
- 2. Build public understanding of human involvement in AI design and application.** People tend to miss both the role of humans in shaping the data that go into AI as well as how AI influences human decision-making, instead assuming AI does everything on its own and functions separately from human intelligence. The challenge here is to reframe the notion of who is behind, who is involved in, and who is affected by AI so that people can contextualize the importance of how it is developed and used.
- 3. Build public understanding of the role of bias in AI.** The public has a limited understanding of how data created by humans, and societal biases, factor into the predictions made by AI. For example, use of AI in policing can result in over-policing communities of color when the system

is trained on arrest data in which those same communities are overrepresented. The challenge here is to reframe the notion of how AI functions in society—how the biases it reflects can do harm to society, communities, and individuals and how data could be used differently.

- 4. Build support for open and transparent development, usage, and regulation of AI and for the fact that AI for the public interest is necessary.** There is a public understanding about what AI is, and the role of humans in creating it, that makes it difficult to think about how to solve the issues surrounding AI. As governments begin to debate and implement AI governance, the public has a limited understanding of why regulation matters. This includes the need to address underlying sources of bias that influence predictive algorithms, the need to increase the oversight and transparency of AI, and the need to diversify the AI field. The challenge here is to reframe what can be done so AI does not cause harm and is used on behalf of the public interest.

We developed a variety of potential frames to address these target ideas and, through qualitative and quantitative research, found an effective and coherent framing strategy.

The Framing Strategy

We have identified a framing strategy that can help productively convey how AI systems replicate social systems of power and harm and can help call for just solutions. In this section we outline specific framing recommendations to achieve this strategy, organized into four types: explanation, metaphor, issues, and values. These frames can be adapted and used in a variety of communications and across specific contexts.

Four Types of Frames

#1: Explanatory Frames

Explanation is a uniquely powerful tool for building people's understanding of social issues. An explanation does not merely assert a relationship between a cause and its effect but shows how one leads to the other. It illuminates process, makes mechanisms visible, and clarifies connections. It helps people recognize injustice and embrace meaningful approaches to solving social problems. Although it differs from definition and description, explanation can incorporate aspects of both framing types, and the explanations we present in this section contain both.

RECOMMENDATION:

Define AI as a technology that, at its core, is a massive data-computation process.

What to do: Use a simple, consistent, and accurate definition of AI technology as a computer process at its core that is trained to find patterns in massive amounts of data.

What this explanation accomplishes: Our quantitative and qualitative analysis suggests that how AI is defined at the most basic level can significantly shift understanding not only of what AI involves, but also of the role of bias in AI and how it can be used for the public interest.

- In our preliminary quantitative testing, this definition of AI led people to disagree with some common misconceptions about AI, such as the idea that AI is always unbiased and accurate and the idea that it is developing consciousness.
- In focus groups, after being shown the definition, participants began to talk about AI as a computer technology process that required human involvement rather than as something objective or autonomous. This seemed to give participants a sense that humans could in fact determine the best use of AI technology.

Why it works: Using the simple definition helps to establish an accurate baseline for a more productive conversation about AI and provides helpful grounding for people when presented with fantastical ideas about what AI is. This definition works because it helps to correct the idea that AI is an intelligent being that poses an existential threat and instead helps people think of AI as data-driven technology that humans can control. Our quantitative

research showed that a definition on its own doesn't seem to be enough to get participants to understand that AI can be biased, but this definition does open the door to a conversation about the role of biased data because it establishes the important premise that AI technology is reliant on data.

We recognize that this definition of AI does not capture every technical nuance. It is not intended to be exhaustive but to provide a clear foundation that allows lay audiences to grasp, at a basic level, what AI is and what it does. Establishing this shared starting point makes it easier for the field to introduce greater complexity over time. Because our goal was to frame the social implications of AI, not to offer a technical primer, this accessible definition proved effective, a finding which is supported by our quantitative research.

Example: Establishing what AI does and how it does it.

AI is a computer process that is used for identifying patterns in large amounts of data.

RECOMMENDATION:

Define AI technology as something different from human intelligence.

What to do: Be explicit about AI not having human awareness or consciousness, but always introduce the idea after having already defined AI as massive data processing, not before and not separately. Emphasize that AI is a technology and is different from human intelligence, with different capabilities; that is, it can process information at a level no human could ever achieve, but it cannot feel emotion or experience the natural world.

What this explanation accomplishes: This explanation productively reframes the comparisons of AI to human intelligence by refocusing the capabilities of AI technology as distinct from what humans are capable of.

- In preliminary quantitative testing, this definition had the highest proportion of participants who claimed that their understanding of AI had changed as a result of reading it.
- In preliminary quantitative testing as well as in the focus groups, this frame reduced the focus on AI being, or becoming, sentient or conscious. This then opens people up to the role that humans play in creating and using AI and to the possibility that the public can play a role in how AI is used.
- In the focus groups and the pre/post interviews, this immediately shifted participants' thinking from viewing AI as an external force that is taking over human society to viewing AI as a technology that humans can control. Participants who, before exposure, argued that AI thinks on its own and could one day "take over" human societies were seemingly reassured by the statement that AI does not have a consciousness, which abated fears concerning AI subjugation of humans. This shifted the conversation away from unrealized dystopian concerns about AI.

Why it works: The concept that AI does not operate with human intelligence and awareness ultimately helps people understand that humans guide the design, purpose, and application of AI. This works hand in hand with the definition of AI as a massive data-computation process and should be used primarily, if not exclusively, as an addition or modifier to that more technical definition. This helps ground people's understanding of AI as a technology, or a tool, designed by humans, which in turn allows them to separate it from human intelligence and all the existential fears that come with that misconception. With their existential fears pushed to the background, people can focus on what they can control about AI in the societies in which they live.

Although it can reduce existential fears, one potential drawback of this frame is that it might reinforce the idea that AI is objective and error free precisely because it uses data and does not possess the fallibility of humans. For this reason, we suggest using this definition in combination with other frame elements that establish how humans are involved in AI and how data can be biased.

Example: Establishing what AI is and what AI is not.

AI is a computer program used to find patterns in large amounts of data, but it isn't aware or conscious like humans are.

RECOMMENDATION:

Show the role of systemic bias in AI by explaining how the data used in AI technology is socially produced.

What to do: Emphasize that AI operates from a massive collection of data, not a given individual's data. Give clear examples of how humans collectively produce data on a daily basis. Focus on how any given set of data can reflect a bias that exists in society because that bias is built into the data that were collected.

What this explanation accomplishes: This explanation provides a way for people to contextualize AI technology in relation to social biases with which they are familiar.

- In survey experiments, an explanation that foregrounded AI processes as data inputs and outputs helped to dispel the idea that AI was unbiased.
- In survey experiments and pre/post interviews, explaining how data are socially produced improved understanding that biased AI data and results can cause harm to low-income populations and communities of color.

Why it works: Framing data as socially produced helps people to apply their understanding of AI as a massive computational process to their social context. Setting the technology in context helps establish the societal nature of the bias in AI. AI does not operate on only one person's data; it operates on a collective set of data from thousands to millions of individuals and groups. Therefore, the bias in the data is a social bias, not an individual one. Once this social bias is established, the data used for AI processes can then be understood to be biased

in the same way that a person might see biases in a variety of social contexts. Ultimately, this gives people a foundation to see the social importance of the data used for AI and the biases it carries, which disproportionately impact communities of color and low-income communities.

Example #1: Establishing that biased data used in AI reflect social biases.

AI computer processes replicate and intensify biases in the data on which it is trained. We live in a biased society, which means data used to train AI can be biased. And because these biases are in the AI's design, they're often echoed in AI results.

Example #2: Focusing on the social importance of biases in AI.

People develop AI from data that contain the biases of society. This means biases in AI can cause real harm to people, especially communities of color. We need to make sure that AI is used in ways that do not further harm historically oppressed communities. AI must be used to advance justice and benefits for all.

RECOMMENDATION:

Describe the extent of human involvement at every stage in the development of AI.

What to do: Explain how people select training data, scrub data, program and test algorithms, and interpret AI-generated results. Explain that people determine the purpose of using AI and that these purposes involve and affect other people.

What this explanation accomplishes: Foregrounding human involvement at every level of the development and application of AI significantly reduces the misconception that AI doesn't need humans to function.

- In survey experiments, explaining human involvement in AI helped to shift people away from the idea that AI does not need humans to function.
- Compared to their initial understanding of AI in the pre/post interviews, this explanation helped participants more fully understand that the processes of AI systems continually involve human input, not only from developers and not just at the initial stage of AI development.

Why it works: In public thinking, AI technology is often positioned as something that operates objectively and autonomously. People tend to miss both the role of humans in shaping the data that go into AI as well as how AI influences human decision-making, instead assuming AI does everything on its own and functions separately from human intelligence. Here, the focus is on AI as a technological process that continually involves human input. With this focus, the explanation helps to promote the agency that people do have in AI processes and outcomes as well as how social biases are reproduced through human involvement in AI technology.

Example #1: Describing the different stages of human involvement in AI development and usage.

Behind every AI process, there are humans responsible for making it work. People choose the data to train the AI, and others clean and process that data before training begins. There are more people who design and program and others who ensure the AI is working as designed. And, eventually, humans are required to interpret AI results and put them to use in the real world.

Example #2: Describing how the human involvement in AI is connected to the problems of bias in AI.

AI doesn't work without the people who design, test, and use it, which means that our social biases are embedded in AI. We need policies to make sure that AI technology is developed in ways that benefit society and can't cause harm.

#2: Metaphor

Metaphors help people think about an abstract, unfamiliar, or misunderstood system or process by comparing it to something familiar and concrete. Metaphors bring people's everyday knowledge to the task of rethinking complex social issues. In this way, metaphors are not only literary devices but also devices for thinking. They can put a picture in the public's mind where none existed before, and they can reshape and update our shared mental images of social issues. When we use them wisely in our communication related to social change, we can amplify their impact.

RECOMMENDATION:

Use the natural language metaphor of an “echo” to talk about how AI data can reflect biases and power in our society.

What to do: Explicitly use the metaphor to explain how AI reflects social and systemic biases rather than the personal biases of individuals. Deliberately use “echo” as a verb when describing what AI does; for instance, “AI echoes what is in the data used to train it.”

What this frame accomplishes: The echo metaphor does an effective job of communicating that the biases in AI are a delayed repetition of the biases in the data, which reflect the biases in society more broadly.

- In the experiments, when the echo metaphor tied the idea of bias to systems of social power, it improved people's understanding that the use of AI can cause harm.

Why it works: The echo metaphor can easily help to structure thinking about AI because it is a natural language metaphor that already is in use in common language; for example, “the sentiment was echoed back to me,” “echoes of the past.” Because the echo metaphor is based on the idea of reflecting back a source of sound, it helps people connect to the idea of social biases in data systems as a source that is reflected in AI systems. This opens the door for people to connect how biases are embedded in AI systems to the ways they can do harm to people and communities.

Advocates, experts, and communications professionals must, however, take care that the “echo” is explained as an echo of social bias. If AI is interpreted as echoing individual bias, this could cause people to place responsibility for bias in AI on individual programmers instead of on systemic biases that are embedded in the data used for AI. As we will see in later recommendations, the echo metaphor is most effective when coupled with a frame of social or systemic power.

Example #1: Echo.

We live in a biased society. This means that any data used to train AI will be biased, and those biases will be echoed in AI results.

Example #2: Echo + Power.

We live in a society where people with power have most of the control, while many others have very little. The data used to create AI is often biased in ways that benefit those in power. Because of this, AI can echo those biases and make social problems even worse.

RECOMMENDATION:

Use an “amplification” metaphor to talk about how AI technology can increase the harm done to already marginalized communities in our society.

What to do: Talk about how the very ability of AI technology to identify, analyze, synthesize, and automate at remarkably high speeds has the effect of reinforcing at scale—amplifying—the systemic biases in the data it uses. Use this metaphor alongside the echo metaphor to explain that if AI echoes back social biases that are in the data, then the proliferation of AI use will help make the echo louder and will extend the reach of the biases.

What this frame accomplishes: Like the echo metaphor, the amplification metaphor uses a common and easily understood natural metaphor to describe what the effects of AI use are on a social level.

- In the pre/post interviews, participants were able to map their understanding of an amplifier (as boosting the sound that passes through it) onto AI as a technology (as boosting the biases in the data that pass through it).
- In the survey experiments, the amplification metaphor improved participants’ understanding that AI is not inherently objective and accurate. It also marginally increased the understanding that biased AI processes can cause harm to people of color.
- In the pre/post interviews, the frame helped participants identify solutions to AI bias. Participants talked about more government regulation, increased training for programmers, media literacy training for general public AI users, and transparency notices for AI that uses biased information.

Why it works: This natural language metaphor is easily available to people and used frequently in everyday language. This makes it easy to use when talking about the effects of AI on society, specifically about the negative effects of the bias in AI systems. This metaphor is fundamentally the idea that a source input is being amplified through a technology. It allows people to scale up the effect that AI technology has on society. When used to explain how bias in AI does harm, this metaphor helps people see how that harm can be projected into more communities through AI. It also helps them see how AI can intensify the harm done to already marginalized communities.

Example:

When AI is used for making decisions, it amplifies society's biases and can cause real harm.

#3: Issues

Issue frames foreground a particular aspect of the topic at hand and shape how the nature of the problem and its solutions are interpreted. These types of frames set the agenda, establishing what an issue is actually “about” and what its end goal is. Changing an issue frame can dramatically affect public thinking and policy support.

RECOMMENDATION:

Situate the use of AI technology as an issue of social power that creates and reinforces systemic inequalities.

What to do: Name the power dynamics of AI: The powerful in society control and benefit from the development and use of AI technology. Directly state that AI technology—not a household device that uses AI—is leveraged by the powerful in society for their own interests, not the public interest. The “powerful” can include but does not have to be confined to technological corporations. It can also be helpful to connect this frame to the echo metaphor to explain how the biases in AI data and AI-generated results can echo social power. This helps people connect bias to the social power that it carries.

What this frame accomplishes: This frame helps people to think of the benefits of AI in terms of social benefit and social harm.

- In the pre/post testing, a frame that foregrounded systems of power in society helped people think of systemic solutions, such as data transparency and improved industry regulation, to address the harm caused in society, specifically to marginalized communities, by the use of AI.
- In the pre/post testing, a frame that foregrounded the human harm caused by the power that guides AI technology helped people understand that AI bias results from data that contain the biases of social power rather than from the individual biases of a programmer.

Why it works: This frame opens the door for people to come to terms with the power dynamics behind the development and use of AI. Focusing on the deeper systemic power structures that AI is born out of helps people see AI as a component of the social system rather than simply a consumer product used to get things done. Once people understand AI technology as an issue

of power, and of the imbalances of power in our society, they can understand the challenges that AI presents for society, and for marginalized Black, Brown, and low-income communities specifically. This frame works particularly well along with the “echo” metaphor, as it causes people to connect the importance of social power in the data to how those biased data shape how AI is used and interpreted.

Example:

A handful of powerful companies in our society have control over the development and use of data in AI. This means that AI systems echo the biases of the very powerful. That causes harm to our society, especially in Black, Brown, and low-income communities. AI technology needs to be used in ways that empower communities and advance justice.

RECOMMENDATION:

Talk about AI technology as an issue of social justice using real-life contextual examples.

What to do: Talk about the social harm and injustice done to human beings by AI technology in a given context, especially to already marginalized communities. Be specific. Show examples of how the use of specific types of AI data can perpetuate problems rather than solve them. Talk about the specific communities that are affected and why they are negatively affected by the use of AI. Through the examples, explain what types of data are used to train the AI’s calculations in this specific context. It is important to paint a social justice vision for the use of AI technology. Contextual examples can help show people a proactive vision for AI as well, rather than using them only to point out problems. Each context will have its unique challenges, so specific examples are most effective in providing a vision.

What this frame accomplishes: Framing the issue of AI through contextual examples can help people see AI technology as an issue of social justice that has concrete and important impacts on the most marginalized communities in our society.

- In focus groups, the introduction of a social justice frame, through contextual examples, helped participants recognize that AI could be potentially helpful in some contexts, but they unanimously agreed that it should not cause harm.
- In the focus groups, this framing shifted the conversation from AI as an existential threat to one about preventing harm, even if the potential harm in each context was understood differently.
- In the focus groups, participants expressed a need to have experienced human emotion and to be able to empathize to effectively solve these problems in their contexts. This drew attention to the non-sentience of AI and helped people see the limits of how the technology could be applied.
- The frame helped focus group participants arrive at the conclusion that we should be using data differently in AI—for the common good.

Why it works: Framing AI technology as an issue of social justice allows people to see how AI technology reinforces inequalities and injustices faced by marginalized groups. When communicated through a lens of social harm and injustice, it becomes more difficult for people to explain away harm done by AI as “glitches” or inaccuracies in data, and it can be understood as being built into the technology itself. The technology, and therefore also the harm done by the technology, is understood to be social and systemic. This is where the use of contextual examples becomes so powerful.

Examples bring to life the harm done by AI technology that, like the massive amounts of data it uses, are otherwise hidden from the public eye. In fact, because the bias in AI echoes the bias in society, many of the harms caused by AI are harms that already exist, which makes it difficult to tie directly to AI. The ability to visualize the need for justice is critical to people seeing AI as a social justice issue. Similarly, it can help to translate the social power of AI technology into a vision for what can be done to prevent harm and, ultimately, to do good in society.

In this way, this frame helps people locate the problems with AI use, and therefore the solutions they can imagine, for the communities that are most negatively affected by the current systems of power in our society. Ultimately, this frame helps people identify the social harm that can be done by AI, but it also helps them think about how AI as a technology can be employed for public interest.

Example:

Social justice issue framing (general): *People develop AI from data that contain the biases of society. These biases in AI can cause real harm to people, especially in Black, Brown, and low-income communities.*

Social justice issue framing in the context of policing: *Police officers use AI to figure out where crime is likely to happen and where to assign patrol units. The data used by AI include zip codes and historical arrest rates. The problem is, zip codes with high Black populations have historically been over-policed. That means they have higher arrest rates. Using these data means the AI predicts crime, and puts more police, in those same Black communities.*

Social justice issue framing in the context of child welfare: *Social workers use AI to calculate the risk that a child will be placed in foster care. The data used by AI include whether a family gets public benefits, like Medicaid and other programs. Many parents with disabilities receive additional support and services from public benefits programs. AI uses benefits data to calculate whether a child is at risk of being placed in foster care. This often means that a higher rate of kids are taken from parents with disabilities and put in foster care.*

Social justice issue framing in the context of health care: *Health insurance companies use AI to determine risks and the need for patient care. This is based on a patient's past spending on health care. Lower spending equals higher risk, and that means getting denied treatment. However, Black families have historically had lower incomes. This means they're likely to spend less on health care than other groups and are labeled high risk. This excludes Black patients from getting extra care, even though they have some of the highest rates of diabetes and cancer deaths.*

#4: Values

Values are cherished ideals or deep-seated principles that are widely shared across a culture. They invite civic thinking about social issues, helping us all see why we should care about them and what's at stake if we don't. Values are especially effective tools for introducing a topic because they can orient people toward the idea of shared benefits from the start and then help them keep an open mind about the message that follows.

RECOMMENDATION:

Invoke the shared value of justice when explaining how AI could be developed and used for the public interest.

What to do: Emphasize that the purpose of any technology in society is to benefit everyone in a society, not just a select few. Speak directly to a vision in which AI technology is used for the public interest and to provide equal and equitable opportunities for all in society. Establish that in order for people to be treated justly in our society, technologies such as AI must be used to reinforce that fair treatment and must not create or perpetuate harm and injustice.

What this frame accomplishes: This frame taps into people's sense of justice that, regardless of one's identity or background, everyone should receive impartial treatment and have no harm done to them.

- In focus groups, this value frame helped participants understand that technology used in the public interest can improve the quality of life for the people in a community, and therefore, AI technology could be used differently, so that it does no harm and better serves the public interest.
- In the survey experiments, this frame helped people see how biased AI processes and AI-generated results cause harm to marginalized communities, specifically communities of color and low-income communities.
- In the pre/post interviews, this frame element appeared to help people understand systemic bias in AI technology and to think about systemic solutions so AI will work for the public interest.

Why it works: People are more able to effectively argue for using AI for the public interest when it is framed as a shared value. In both the qualitative and quantitative research, participants agreed with the idea that everyone should be treated impartially and not suffer harm. Furthermore, people do not think that technology should violate any of these values. In this way, once established as something that can uphold our shared value of justice—and as something that is currently violating that shared value—AI technology begins to be understood as having important social implications. From this value frame of justice, people are able to recognize past and present injustices and to imagine how AI technology can be used to set us on a path to achieve a just society.

Example:

All people must be treated justly, no matter where they come from, the color of their skin, or how much money they make. We must come together and demand that AI be developed and used to advance justice and to benefit the public good.

Summary of Recommendations

Recommendation	Why?	Example
Explanations & definitions		
Define AI as a massive data-computation process.	Helps ward off misconceptions about AI as being objective, autonomous, or conscious and opens the door for human involvement.	AI is a computer process that is used for identifying patterns in large amounts of data.
Define AI technology as something different from human intelligence.	Helps build the understanding that AI is something humans control rather than a force that will take over the world.	AI is a computer process used to find patterns in large amounts of data, but it isn't aware or conscious like humans are.
Show the role of systemic bias in AI by explaining how data used in AI technology is socially produced.	Helps build understanding that biased AI can harm low-income populations and communities of color.	<p>Example #1:</p> <p>AI computer processes replicate and intensify biases in the data it is trained on. We live in a biased society, which means data used to train AI can be biased. And because these biases are in the AI's design, they're often echoed in AI results.</p> <p>Example #2:</p> <p>People develop AI from data that contain the biases of society. That means that biases in AI can cause real harm to people, especially communities of color. We need to make sure that AI is used in ways that do not further harm historically oppressed communities. AI must be used to advance justice and benefits for all.</p>
Describe the extent of human involvement at every stage in the development of AI.	Builds greater understanding of the human input that goes into AI across different stages of development and usage.	<p>Example #1:</p> <p>Behind every AI process, there are humans responsible for making it work. People choose the data to train the AI, and others clean and process that data before training begins. There are more people who design and program and others who ensure the AI is working as designed. And, eventually, humans are required to interpret AI results and put them to use in the real world.</p> <p>Example #2:</p> <p>AI doesn't work without the people who design, test, and use it, which means that our social biases are embedded in AI. We need policies to make sure that AI technology is developed in ways that benefit society and can't cause harm.</p>

Recommendation	Why?	Example
Metaphor		
Use the natural language metaphor of an “echo” to talk about how AI data can reflect the biases of our society.	Helps connect the idea of bias in AI to social bias.	We live in a biased society. This means that any data used to train AI will be biased, and those biases will be echoed in AI results.
Add the element of “power” to the echo metaphor to talk about the impact of the bias in AI data and development.	Helps connect the idea of bias in AI to the systems of power that create the bias and the consequences of that bias.	We live in a society where people with power have most of the control, while many others have very little. The data used to create AI are often biased in ways that benefit those in power. Because of this, AI can echo those biases and make social problems even worse.
Use an “amplification” metaphor to talk about how AI technology can increase the harm done to already marginalized communities in our society.	Draws people's attention to the ability of AI to quickly and significantly reinforce systemic biases.	When AI is used to make decisions, it amplifies society's biases and can cause real harm.

Issue frames

Situate the use of AI technology as an issue of social power that creates and reinforces systemic disadvantages and inequalities.	Foregrounds how AI causes harm and helps people move toward solutions that address power structures in society.	A handful of powerful companies in our society have control over the development and use of data in AI. This means that AI systems echo the biases of the very powerful. That causes harm through our society, especially in Black, Brown, and low-income communities. AI technology needs to be used in ways that empower communities and advance justice.
Use real-life contextual examples to talk about AI technology as an issue of social justice.	Shifts conversations from AI as an existential threat to the need to design AI in ways that prevent harm.	People develop AI from data that contains the biases of society. These biases in AI can cause real harm to people, especially in Black, Brown, and low-income communities. (See page 17 for more specific examples that can be used in the context of policing, child care, and health insurance).

Values

Evoke the shared value of justice when explaining how AI could be developed and used for the public interest.	Helps people see how technology used for the public interest can improve the quality of life for the people in a community.	All people must be treated justly, no matter where they come from, the color of their skin, or how much money they make. We must come together and demand that AI be developed and used to advance justice and to benefit the public good.
---	---	--

Endnotes

1. *Unleashing the power of how: an explanation declaration.* (2019). FrameWorks Institute. https://www.frameworksinstitute.org/app/uploads/2020/06/fw20_unleash_the_power_of_how.pdf
2. For the survey experiment data that supports this recommendation, see Graph 1 in Appendix F of the Methods Supplement for “Framing the Social Implications of AI.”
3. For the data that supports this recommendation, see Graph 2 in Appendix F of the Methods Supplement for “Framing the Social Implications of AI.”
4. FrameWorks Institute. (2020). *What’s in a frame?* <https://www.frameworksinstitute.org/articles/whats-in-a-frame/>
5. See Graph 3 in Appendix F of the Methods Supplement for “Framing the Social Implications of AI.”
6. See Graph 4 in Appendix F of the Methods Supplement for “Framing the Social Implications of AI.”
7. FrameWorks Institute. (2018). *Framing and policy making.* <https://www.frameworksinstitute.org/articles/framing-and-policy-making/>
8. FrameWorks Institute. (2020). *What’s in a frame?* <https://www.frameworksinstitute.org/articles/whats-in-a-frame/>

Framing the Social Implications of AI

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the FrameWorks Institute.

FrameWorks Institute. (2026). *Framing the Social Implications of AI*. Washington, DC: FrameWorks Institute.

Designed by Constructive · © FrameWorks Institute 2026

