TechTonic Justice

People first. A more just tech future.

SIX ROUNDS OF MESSAGE TESTING

What people actually told us about AI harms

A plain language guide to what was tested, what the numbers mean, and what people are telling us about AI harm.

6	+15.7	+24.7	70-81%
rounds tested	strongest concern	top slogan	guardrail support

What is message testing?

Message testing helps us understand not simply whether people are worried about AI, but **what kinds of harm feel most real, who people hold responsible, where they believe AI should not be used, and what they want done about it**. Across six rounds of research between March and July 2026, TechTonic Justice tested different ways of talking about AI harms, real-world examples, the institutions using these systems, possible rules and protections, and the language people find most meaningful.

The results matter because public debate about AI can easily become abstract. For policymakers, they show where there is public permission for stronger action. For movement organizations, they show how to communicate harms in ways that connect with lived experience. For the wider public, they make visible a key distinction: **people are often less worried about AI itself than about what happens when institutions use it to gain power over consequential parts of their lives.**

How to read the results

MaxDiff	A relative ranking. Positive means an item was chosen as more important more often; negative means it ranked lower against the other options. It is not a percentage.
Concern / support %	A normal percentage: the share saying they were concerned or supported a proposal.
Forced choice	People had to select one top priority even when several things worried them.
Subgroups	Party, age and other differences can be useful, but small subgroups are less precise.

BOTTOM LINE

The public is not simply "anti-AI".

Concern rises when automated systems become gatekeepers to health care, benefits, housing, work or other essentials - especially when people cannot challenge the decision or find a responsible human.

THE BIG PICTURE

What surprised us

Six rounds of testing produced a remarkably consistent picture. People's deepest concerns are not really about the technology in isolation. They are about **power, control, accountability and what can happen when an automated system stands between a person and something they need.**

1 It is not really about "AI." It is about power over people.

The strongest concerns appear when AI is used to **deny, remove, price, monitor or control something important**: health care, benefits, housing, employment or access to a human being. This is deeper than simply asking for a "human in the loop." The question is who should have the power to make consequential decisions about people's lives in the first place.

2 People are willing to draw real red lines.

Support does not stop at transparency statements or voluntary safeguards. Large majorities backed **safety requirements, meaningful penalties, off switches, community authority and outright bans on some high-stakes uses**. Concern can translate into permission for substantial intervention.

3 The strongest messages are concrete, specific and human.

"AI may create bias" is abstract. "Your insurer used AI to deny the treatment your doctor says you need" tells people **who acted, what they did and what was taken away**. A useful communications formula is: **Person + institution + action + consequence.**

4 People want someone accountable.

A faceless algorithm is a weak villain. An insurer denying treatment, an employer monitoring workers, a landlord using automated screening or a contractor cutting off benefits gives people something concrete to understand and challenge. The institution does not have to be universally disliked: **the conflict is often about the use of power.**

5 What people mention first is not necessarily what worries them most.

When asked openly, people readily named **privacy, surveillance, job loss and misinformation**. But once they saw specific examples involving health care, benefits, employment, pricing and housing, those concrete harms generated very high concern. Familiar AI concerns may be easiest to recall without being the deepest source of concern.

6 Bias and discrimination did not win as stand-alone frames.

Bias and discrimination ranked comparatively low when tested against concrete harms. That **does not mean people do not care about discrimination**. It means the stand-alone frame was less compelling than a specific consequence such as denied medical care. The research did not test whether disparity evidence becomes powerful when attached to a concrete harm - an important future question.

7 People want prevention and repair, not just better technology.

"Prevent, Protect, Repair" overwhelmingly beat the alternative phrases, and the final round showed large majorities supporting strong mechanisms to make those principles real. People respond to a **system of responsibility**: prevent harm where possible, protect people while systems are used, and repair harm when prevention fails.

THE BIG LESSON

Do not lead with AI. Lead with people.

Show **who is using the technology, what power it gives them, what a person could lose, and what should happen to prevent or repair the harm.**

N=1,242 LIKELY VOTERS

Main concerns about government use of AI

BOTTOM LINE

Concrete consequences beat abstract warnings.

WHAT THIS ROUND TELLS US

When people compared different concerns about government use of AI, the strongest worries were not technical. People were most concerned that public services would become harder to use, that a single mistake could affect millions, and that AI could be used to mislead the public. The messaging lesson is straightforward: **start with what changes for the person using the service**. An abstract warning about opacity or bias is less immediate than showing someone unable to get help, wrongly losing support, or facing an error nobody will take responsibility for.

+6.3
public services harder to use

+5.3
one mistake harms millions

+3.8
misleading the public

Concern tested	MaxDiff
Public services become harder to use	+6.3
One mistake can harm millions	+5.3
Misleading the public with fake media	+3.8
No one takes responsibility for harm	+2.6
Real wrongful fraud flags / benefit cuts	+2.1
Predicting what people may do	-0.9
Treating mistakes as fraud	-1.7
AI looks objective when it is not	-2.5
Cuts presented as efficiency	-2.9
Selective or unequal enforcement	-3.5
People cannot see how decisions were made	-3.9
Sweeping claim that AI degrades services	-4.7

BOTTOM LINE

Concrete consequences beat abstract warnings.

The strongest message was simple: technology gets between people and the public help they need. Scale also matters - one automated error can affect millions.

N=1,193 LIKELY VOTERS

Which broader AI harms feel most serious?

BOTTOM LINE

The doctor-versus-algorithm conflict is the standout.

WHAT THIS ROUND TELLS US

One result towers over everything else: an insurer using AI to deny doctor-recommended care. The example combines several features that recur throughout the research: **something essential is at stake, a recognizable institution is responsible, a human judgment is being displaced, and the person experiencing the harm has something concrete taken away**. It is also notable that data-center pressure on water and electricity came second, showing that the physical and community consequences of AI infrastructure can resonate when described concretely.

+15.7
insurer denies doctor-recommended care

+5.2
data centres

+3.7
public outsourcing

Harm tested	MaxDiff
Insurer AI denies doctor-recommended care	+15.7
Data centres strain water and power	+5.2
Government outsources programs to tech firms	+3.7
Benefits / housing / jobs cut, no accountability	+1.3
Tracking data without meaningful consent	+0.5
Major decisions cannot be explained	+0.3
Big Tech concentration of power	+0.2
Worker monitoring / scheduling / wage control	-1.0
AI economy leaves struggling people behind	-1.5
Wrongful decisions reduce civic participation	-2.8
Bias and discrimination	-7.7
Creative work used without permission/payment	-13.8

BOTTOM LINE

The doctor-versus-algorithm conflict is the standout.

People react most strongly when AI becomes a gatekeeper between a person and something essential. Concrete institutional action is more powerful than abstract warnings about corporate power.

N=1,167 LIKELY VOTERS

Who do people want controlling AI decisions about their

BOTTOM LINE

The issue is consequential decision power.

WHAT THIS ROUND TELLS US

People are uneasy about giving almost any powerful institution permission to put AI in charge of consequential decisions. Health insurers generated the strongest concern, but landlords, government contractors, state officials, Big Tech companies and others were close behind. Even employers, who were viewed comparatively more favorably, still produced substantial unease. This tells us something important about the "villain" in AI messaging: **the institution does not need to be universally hated**. A person can trust or even like an organization and still believe it should not have certain automated powers over them.

68%
health insurers

67%
landlords / contractors

53%
employer

Actor	At least somewhat upset
Health insurance companies	68%
Donald Trump	67%
Trump administration officials	67%
ICE	67%
Landlords	67%
Private companies running public programs	67%
State government officials	66%
Elon Musk	65%
Big Tech companies	64%
Employer	53%

Intensity vs breadth

Named political figures can produce very intense reactions but also polarise. Everyday gatekeepers create broader shared unease.

Trust vs permission

Employers were generally viewed more favourably, yet 53% were still at least somewhat upset about employer-controlled AI.

BOTTOM LINE

The issue is consequential decision power.

Almost nobody gets a free pass to put AI "in charge" of decisions about people's lives. The important question is who has permission to use it, and for what.

N=1,164 LIKELY VOTERS

Which specific uses of AI cross the line?

BOTTOM LINE

Specific uses reveal a broad boundary.

WHAT THIS ROUND TELLS US

This round exposes the gap between **what people immediately associate with AI** and **what becomes most concerning once they see what AI can actually be used to do**. Without examples, privacy, surveillance, job loss and misinformation came readily to mind. But when people saw specific scenarios, concern was extremely high across health care, disability and Social Security, benefits, workplace monitoring, pricing, housing and being forced to deal with an AI instead of a person. The communications implication is significant: **do not assume the most familiar AI issue is the strongest one. Give people concrete examples of automated power in everyday life.**

73%
care denial concern

27%
#1: Social Security / disability

897
open-ended responses

Concrete AI use	Ext./very concerned	#1 priority
Doctor-recommended treatment denied	73%	11%
Social Security / disability cut off	72%	27%
Falsely accused of benefits fraud	71%	7%
Charged more for groceries	71%	3%
Job rejection with no human involved	70%	5%
Everything at work monitored	70%	7%
Forced to deal with AI instead of a person	69%	6%
Medicaid cut off	68%	6%
Spied on at home to build eviction case	68%	6%
Transplant / pain medicine decision	68%	3%
Rent raised using automated pricing	67%	5%
Benefits repayment demand	65%	2%
Apartment rejected by faulty background check	65%	1%
Child falsely flagged as school threat	64%	5%
Home or car loan denied	62%	2%

Write-ins before any frame: privacy/data/surveillance 20% • job loss/economic harm 16% • misinformation/deepfakes 10%

BOTTOM LINE

Specific uses reveal a broad boundary.

People may associate AI spontaneously with privacy, jobs or deepfakes, but concrete scenarios show strong concern about benefits, care, surveillance, pricing and decisions with no human involved.

N=1,219 LIKELY VOTERS

Which three-word response to AI harm feels right?

BOTTOM LINE

"Prevent, Protect, Repair" gives TTJ a lifecycle of responsibility.

WHAT THIS ROUND TELLS US

This round asked a deceptively simple question: what language best describes what should happen when AI can cause harm? "Prevent, Protect, Repair" won overwhelmingly. The individual words matter. **Prevent** says harm should not simply be accepted as inevitable. **Protect** puts the emphasis on people while systems are being used. **Repair** says responsibility continues after something goes wrong and the harm should actually be put right. Importantly, the phrase also performed strongly across partisan groups.

CLEAR WINNER

Prevent, Protect, Repair

+24.7 MaxDiff

Prevent before harm • Protect while systems are used • Repair when prevention fails

Phrase tested	MaxDiff
Prevent, Protect, Repair	+24.7
Stop, Protect, Repair	+7.1
Prevent, Stop, Fix	+2.7
Stop, Protect, Fix	-4.3
Prevent, Stop, Treat	-7.5
Prevent, Mitigate, Repair	-22.6

What the wording tells us

Prevent beats Stop. Protect beats Mitigate by a huge margin. Repair beats Fix. The verbs are active, human and easy to understand.

Cross-partisan

The winning phrase performed strongly among Democrats, Independents and Republicans.

BOTTOM LINE

"Prevent, Protect, Repair" gives TTJ a lifecycle of responsibility.

It moves the conversation from vague "responsible AI" to clear duties before, during and after harm.

N=1,058 LIKELY VOTERS

What rules are people actually willing to support?

BOTTOM LINE

Concern has become permission for meaningful intervention.

WHAT THIS ROUND TELLS US

The final round answers one of the biggest questions raised by the previous research: **does concern actually translate into support for action?** The answer is yes. Every major guardrail tested received support of 70 percent or more, including safety proof, off switches, meaningful penalties, community authority and bans on some high-stakes uses. More people worried that regulation would **not go far enough** than worried about excessive intervention. Policymakers should be careful about assuming that strong AI rules are beyond what the public will accept.

72%
support full package

32%
worry rules do not go far enough

48%
maker + user equally responsible

Policy proposal	Support
Meaningful penalties	77-81%
Off switches	79-80%
Safety proof before release	79-80%
High-stakes bans	70-80%
Safety proof before use	77%
Community boards with real/legal authority	74-75%

More detail

High-stakes bans

Bans on AI decisions that harm or deny health care, housing or work were the top priority in both split samples.

Burden of proof

People support requiring companies and institutions to prove systems are safe before release or use.

BOTTOM LINE

Concern has become permission for meaningful intervention.

Large majorities support proof of safety, off switches, penalties, community authority and bans. Voters are more worried rules will be too weak than too strong.