# Against Trust in Technology

Maggie Oates
Carnegie Mellon University

# What kind of ‘trust’ I’m talking about:

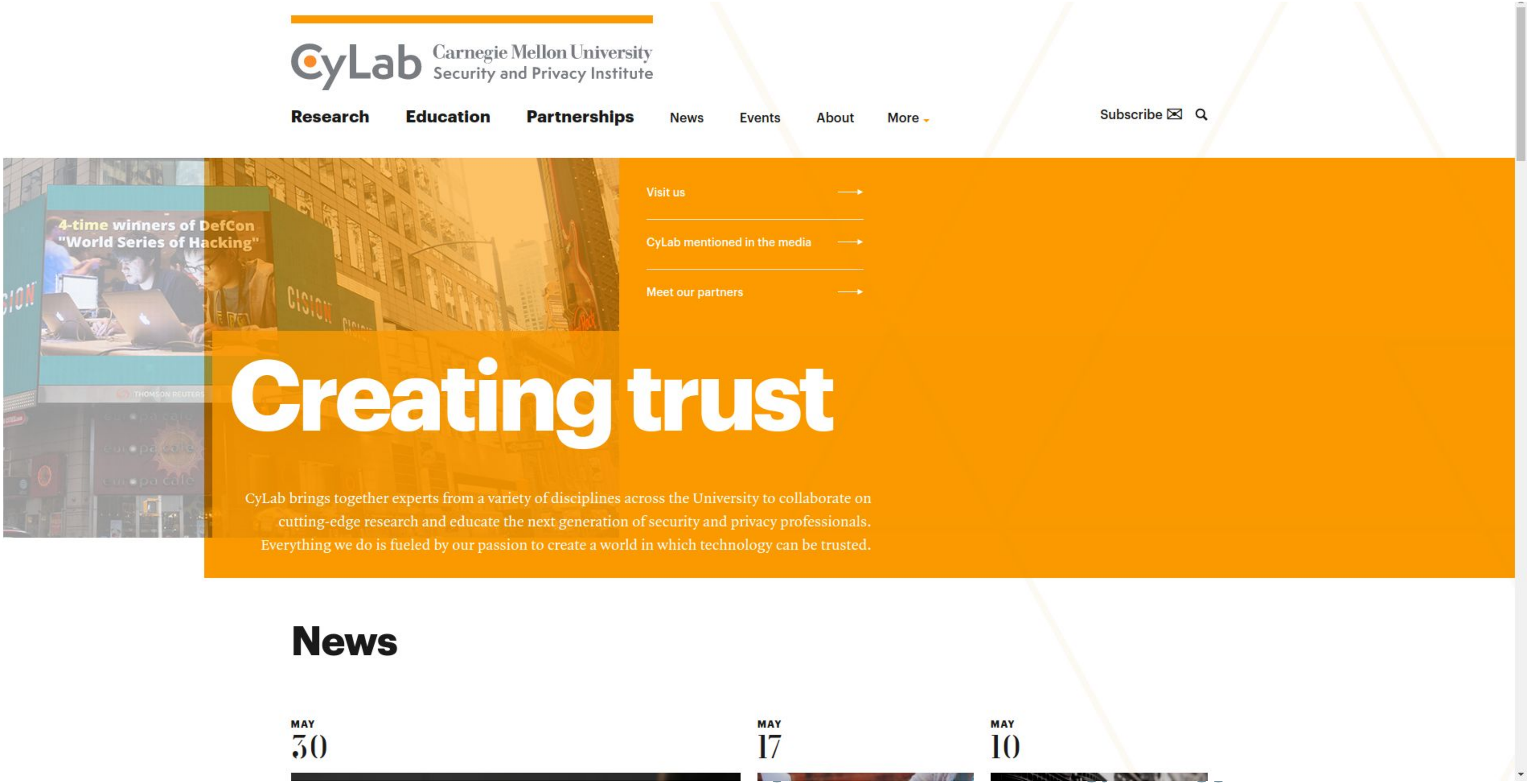
❌ Cryptographic trust

❌ Trust in a company

❌ Trust in peer-to-peer services, sharing economy

✅ Trust in technological artifacts

✅ Trust in technology as a research construct



# Creating trust

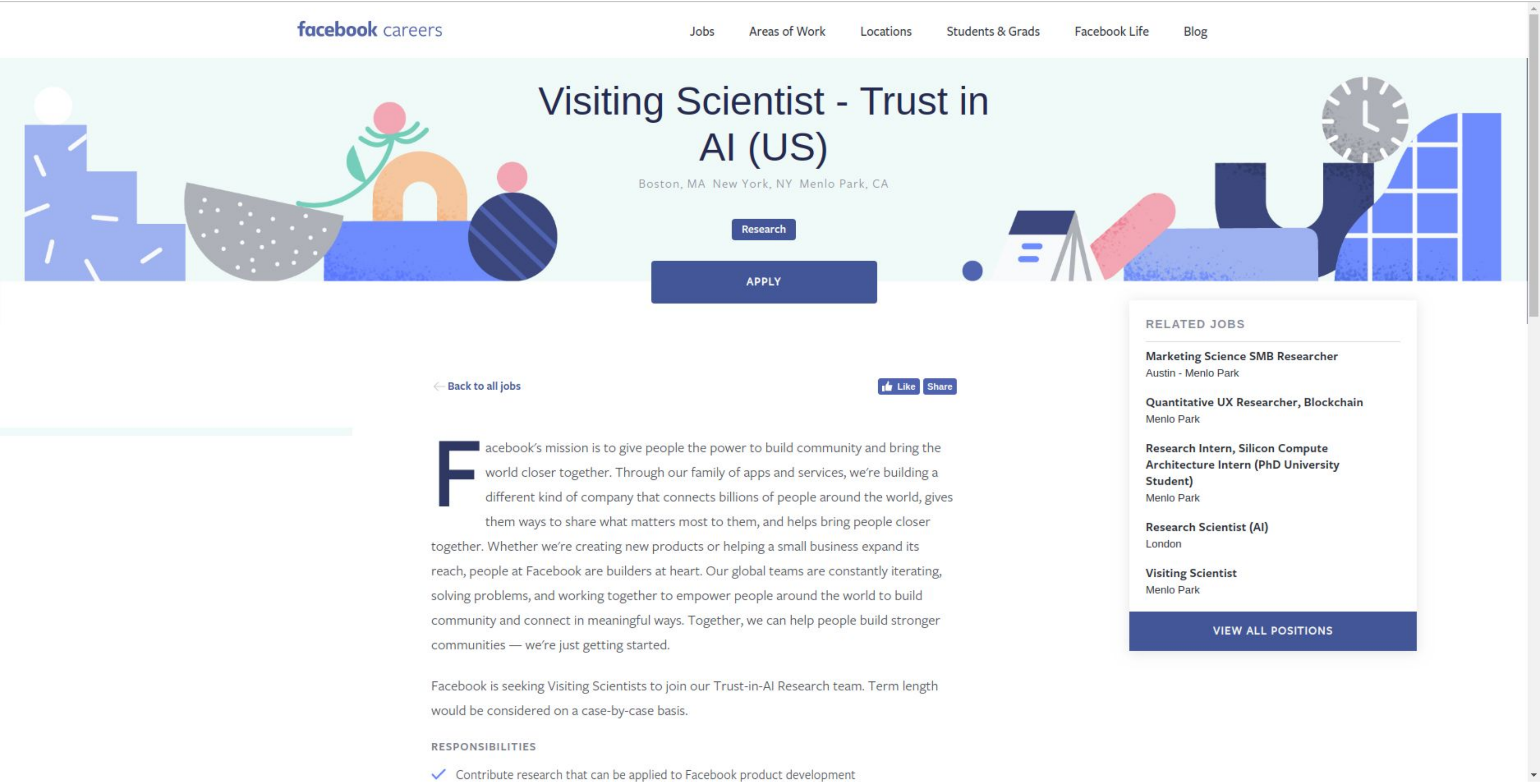
CyLab brings together experts from a variety of disciplines across the University to collaborate on cutting-edge research and educate the next generation of security and privacy professionals. Everything we do is fueled by our passion to create a world in which technology can be trusted.

## News

MAY 30

MAY 17

MAY 10



# Visiting Scientist - Trust in AI (US)

Boston, MA New York, NY Menlo Park, CA

Facebook's mission is to give people the power to build community and bring the world closer together. Through our family of apps and services, we're building a different kind of company that connects billions of people around the world, gives them ways to share what matters most to them, and helps bring people closer together. Whether we're creating new products or helping a small business expand its reach, people at Facebook are builders at heart. Our global teams are constantly iterating, solving problems, and working together to empower people around the world to build community and connect in meaningful ways. Together, we can help people build stronger communities — we're just getting started.

Facebook is seeking Visiting Scientists to join our Trust-in-AI Research team. Term length would be considered on a case-by-case basis.

RESPONSIBILITIES

✓ Contribute research that can be applied to Facebook product development

RELATED JOBS

**Marketing Science SMB Researcher**
Austin - Menlo Park

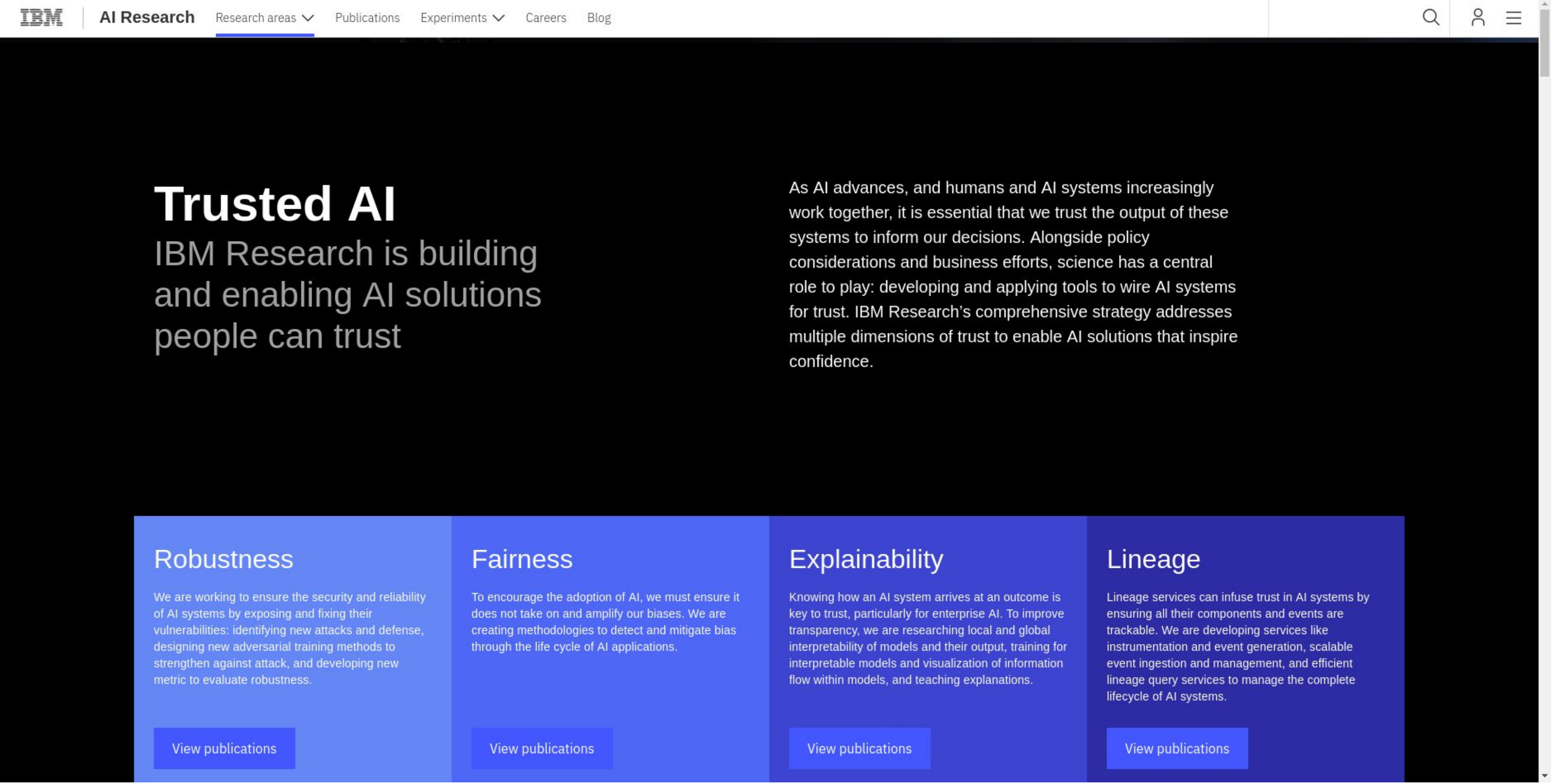
**Quantitative UX Researcher, Blockchain**
Menlo Park

**Research Intern, Silicon Compute Architecture Intern (PhD University Student)**
Menlo Park

**Research Scientist (AI)**
London

**Visiting Scientist**
Menlo Park

"Against Trust in Technology" - Maggie Oates - SOUPS 19



# Trusted AI

## IBM Research is building and enabling AI solutions people can trust

As AI advances, and humans and AI systems increasingly work together, it is essential that we trust the output of these systems to inform our decisions. Alongside policy considerations and business efforts, science has a central role to play: developing and applying tools to wire AI systems for trust. IBM Research's comprehensive strategy addresses multiple dimensions of trust to enable AI solutions that inspire confidence.

### Robustness

We are working to ensure the security and reliability of AI systems by exposing and fixing their vulnerabilities: identifying new attacks and defense, designing new adversarial training methods to strengthen against attack, and developing new metric to evaluate robustness.

### Fairness

To encourage the adoption of AI, we must ensure it does not take on and amplify our biases. We are creating methodologies to detect and mitigate bias through the life cycle of AI applications.

### Explainability

Knowing how an AI system arrives at an outcome is key to trust, particularly for enterprise AI. To improve transparency, we are researching local and global interpretability of models and their output, training for interpretable models and visualization of information flow within models, and teaching explanations.

### Lineage

Lineage services can infuse trust in AI systems by ensuring all their components and events are trackable. We are developing services like instrumentation and event generation, scalable event ingestion and management, and efficient lineage query services to manage the complete lifecycle of AI systems.

“If people don’t trust this tech, they won’t use it.”

“Against Trust in Technology” - Maggie Oates - SOUPS 19

End goal:

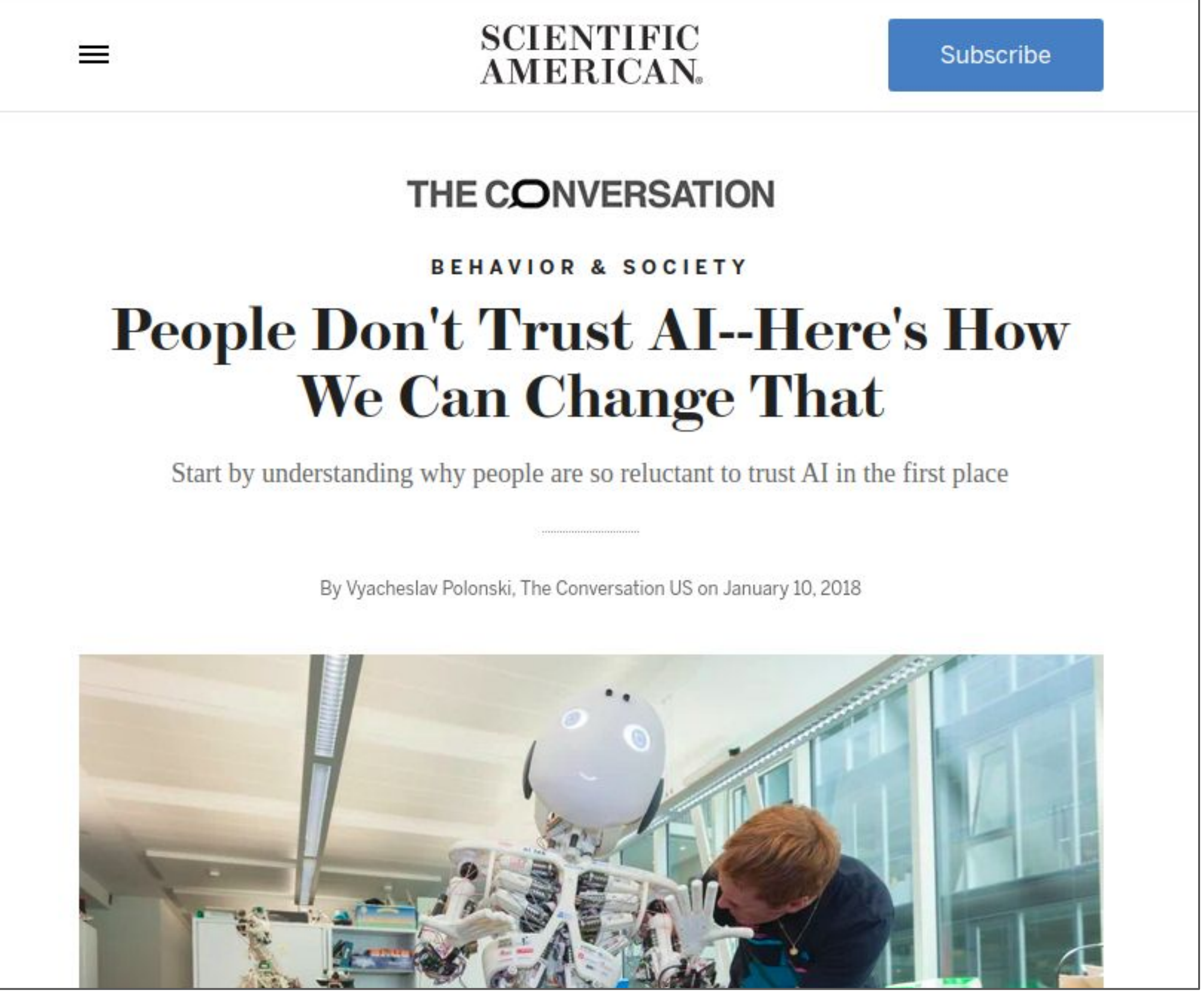
Design & build
Desirable technology

1. ‘Trust’ is an experience of the user, not a property of the artifact
2. “Building trust” is largely a job for marketers, not engineers
3. *If* there is Desirable tech that people refuse to use, then think of it as an Adoption problem, not a trust problem

“Against Trust in Technology” - Maggie Oates - SOUPS 19

Trust is a likely side-effect of Desirable tech, *not an end goal*.

“Against Trust in Technology” - Maggie Oates - SOUPS 19



SCIENTIFIC AMERICAN

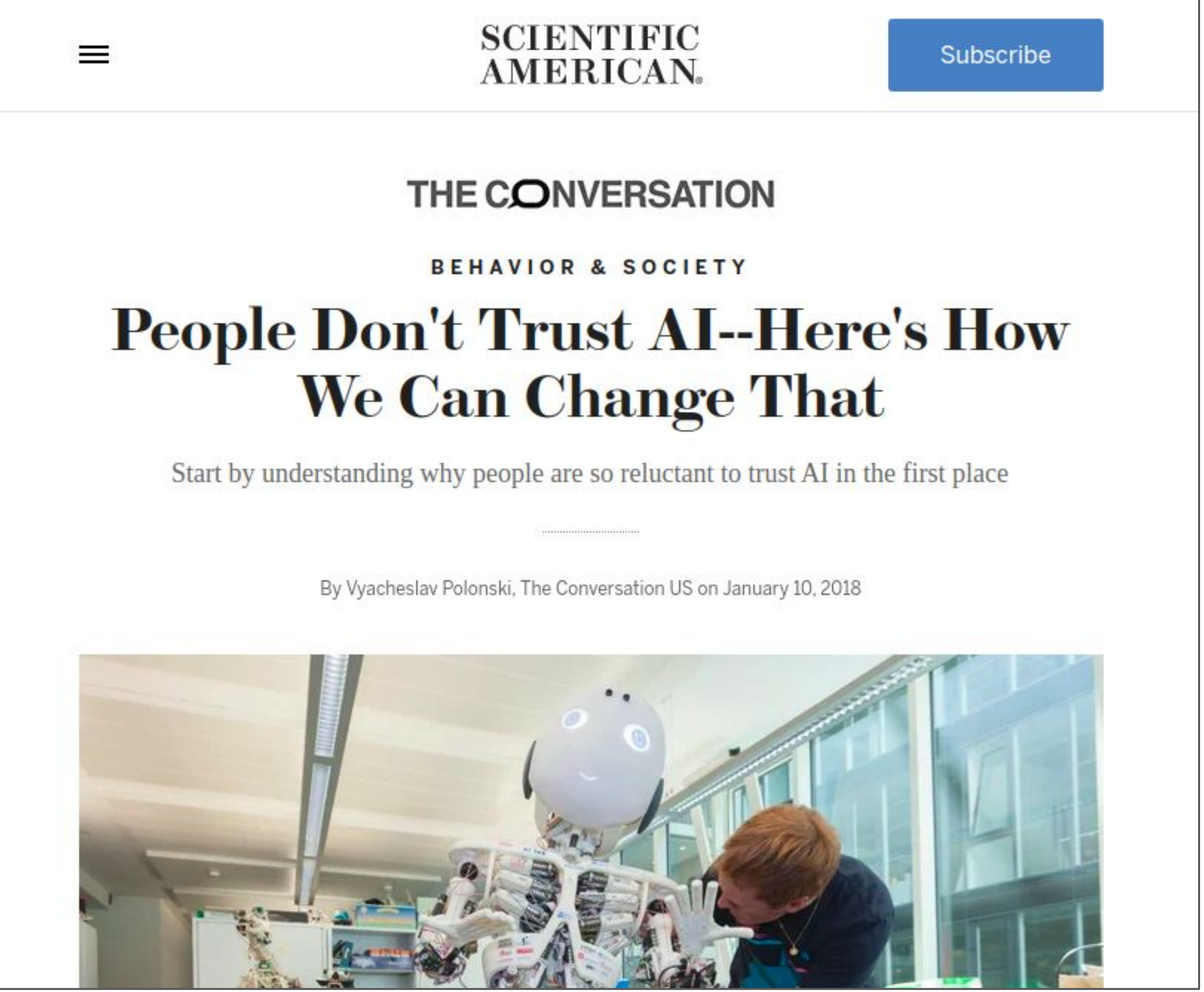
THE CONVERSATION

BEHAVIOR & SOCIETY

# People Don't Trust AI--Here's How We Can Change That

Start by understanding why people are so reluctant to trust AI in the first place

By Vyacheslav Polonski, The Conversation US on January 10, 2018



SCIENTIFIC AMERICAN

THE CONVERSATION

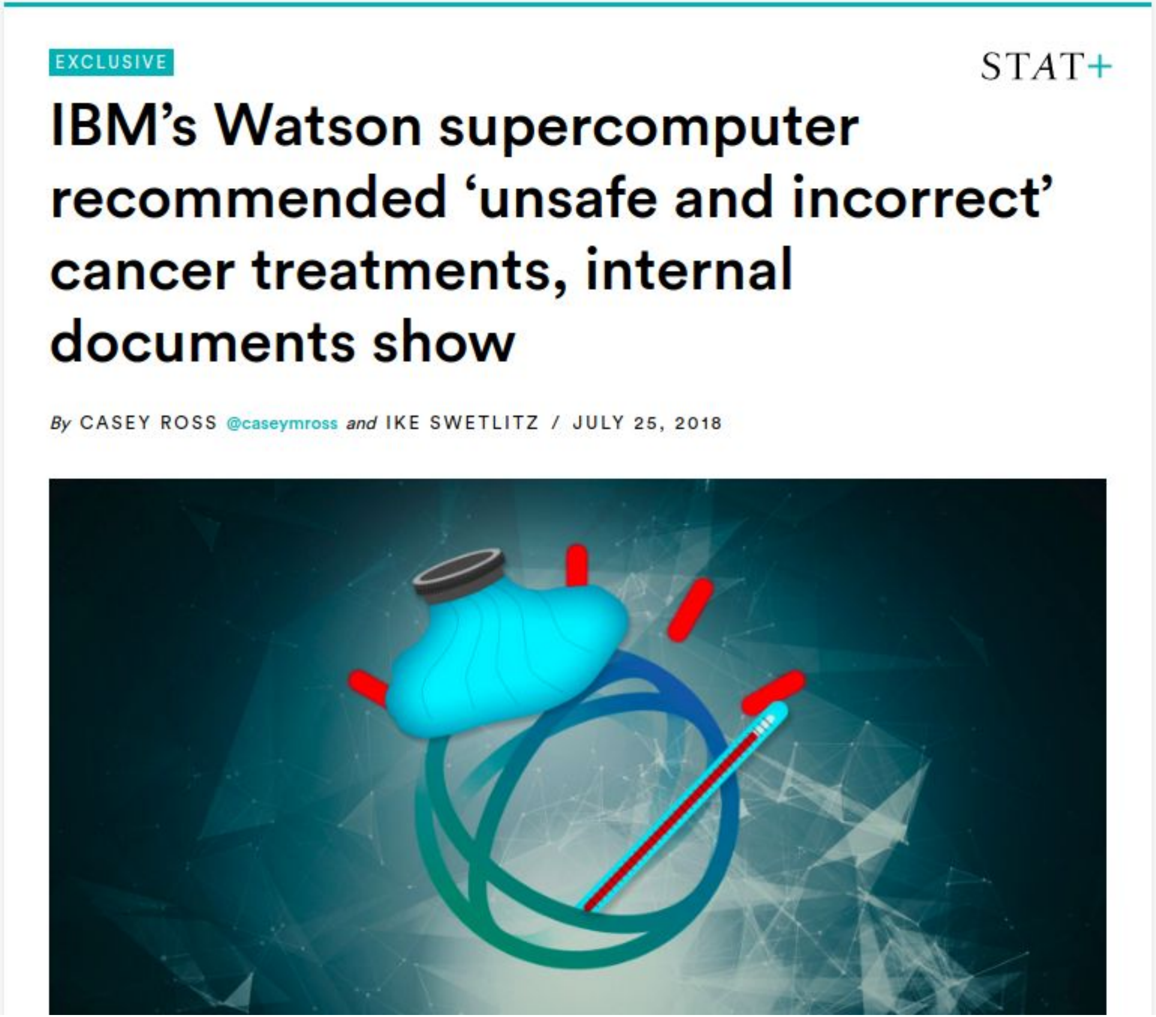
BEHAVIOR & SOCIETY

# People Don't Trust AI--Here's How We Can Change That

Start by understanding why people are so reluctant to trust AI in the first place

By Vyacheslav Polonski, The Conversation US on January 10, 2018

“The problem with Watson for Oncology was that doctors simply didn’t trust it”



EXCLUSIVE

STAT+

# IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show

*By* CASEY ROSS @caseymross *and* IKE SWETLITZ / JULY 25, 2018

"The problem with Watson for Oncology was that doctors simply didn't trust it"

Don’t build for trust, build with values.

Don’t measure trust, 📏 measure…(whatever it is you *actually* want to measure)

# Instead of trust?

- Describing Desirable tech
  - Trustworthiness
  - Contextual integrity
  - Value-aligned design
- Adoption and user behavior
  - Rhetorically transparent models of adoption (e.g., UTAUT, TAMS)
- User perception
  - Psychological safety
  - Perceived Risk
  - Emotional valence

Research is not merely a scientific exercise; it is also a rhetorical one.