# Integrity Institute
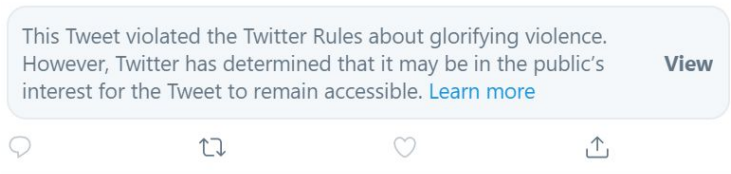
Pitch Deck

The social internet should help
individuals, society, and democracy thrive

# Instead, we're seeing...



This Tweet violated the Twitter Rules about glorifying violence. However, Twitter has determined that it may be in the public's interest for the Tweet to remain accessible. Learn more

View

**The Washington Post**
*Democracy Dies in Darkness*

Sections

THE TECHNOLOGY 202 • Analysis

## Facebook's most popular posts show how it's vulnerable to exploitation, report finds

By Cristiano Lima

March 30, 2022 at 9:06 a.m. EDT

**BLEEPINGCOMPUTER**

NEWS ▾   DOWNLOADS ▾   VIRUS REMOVAL GUIDES ▾   TUTORIALS

**Twitter bots pose as support staff to steal your cryptocurrency**

By Lawrence Abrams                                     December 7, 2021

▷ Harm to individuals: Spam, hoaxes, bots, harassment, hate speech, misinformation, disinformation, and other harms

▷ Bad behavior being **structurally rewarded** by platforms - companies can use bad design practices or fail to build responsibly.

▷ **Harms to societies**: viral lynchings in India; incitement of genocide in Myanmar; civil violence in Ethiopia

▷ **Community-wide impact**: even if you abstain from using a social media platform, your neighbors, family members, and fellow voters still do

# Many groups try to address this problem, but...

**THEY ALL LACK EXPERIENCE ACTUALLY WORKING ON PLATFORMS**

This is important because people with experience can:

▷ **Credibly explain** what approaches actually work
  ▷ Educate stakeholders on issues and how to fix them
    ○ Policymakers, advocacy groups, NGOs, journalists, and academics
  ▷ Help companies decide what to build and how to do it
  ▷ Share best practices around technical solutions
▷ **Credibly rebut** bad-faith arguments
  ▷ "Company X is bad because they did Y"
  ▷ "Proposed regulation Z is impossible to implement, and thus shouldn't be adopted"
▷ Frame the problem correctly and **avoid rabbitholed arguments**
  ▷ "Censorship vs. Free Speech" is a distraction, and we can tell you why
  ▷ "This isn't actually hate speech" vs. "Yes, it is" is not relevant

# Integrity Professionals Are The Key

**Integrity Professionals** are tech workers with experience addressing and mitigating harms to people and societies within social Internet platforms.

Integrity Professionals have experience in tackling these and other issues on behalf of social Internet companies:

- ▷ Ethical Design
- ▷ Harassment
- ▷ Hate Speech
- ▷ Hoaxes
- ▷ Spam
- ▷ Inauthentic Behavior
- ▷ Data Transparency and Reporting
- ▷ Information Operations
- ▷ Content Quality
- ▷ Espionage

- ▷ Disinformation
- ▷ Misinformation
- ▷ Digital Crime
- ▷ Information Ecosystems
- ▷ Child Safety
- ▷ Counter-terrorism
- ▷ Human Trafficking
- ▷ Toxicity
- ▷ Impersonation
- ▷ … and more!

This expertise was locked inside a few big companies until the Integrity Institute

Our growing community has 78 members with experience across 20+ platforms (including all the companies seen here).

# The Integrity Institute brings on-platform experience to the people theorizing, building, and governing the social Internet

▷ Diversity in experience: **over 70 active members** spanning **more than 18 platforms**

▷ Ethics-first: **non-profit 501(c)(3)**, foundation-funded

▷ Breadth of knowledge: active global contributions **across the Integrity spectrum** - developing the theory of Integrity, advising on effective legislation, advising on technical implementation, monitoring for compliance, and more

**ON OMIDYAR NETWORK**

*"We are proud to support the Integrity Institute's efforts to shape a better social internet. The organization and its network has already **achieved so much in a short period of time, and is key to our collective ability to hold platforms accountable to the public interest**."*
– Anamitra Deb, Omidyar Network

**THE WALL STREET JOURNAL**

*"The institute is now advising **lawmakers and think tanks around the world**"*

**The New York Times**

*2021 Good Tech Award*

# We Bridge Integrity Professionals To The World



**Tech Companies**
▷ How do we build a platform that will work over a long period of time?
▷ How do we stop harms before PR crises?
▷ How do we get credit when we do well?
▷ What are easy solutions that work?

**Policymakers and NGOs**
▷ How do platforms work?
▷ What regulation would have a positive impact on business incentives?
▷ What are the real problems to solve?

**Other Integrity Workers**
▷ Has anyone worked on this problem before?
▷ How do I convince my company to prioritize my team?
▷ How do we get regulation that helps us do our job?

**Academia and Researchers**
▷ Are we studying the right issues?
▷ Are we using the right data?
▷ How do you study platforms from the inside?

# Our Team

**Jeff Allen**
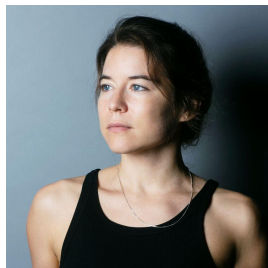Co-founder, Chief Research Officer

**Sahar Massachi**
Co-founder, Executive Director

**Rachel Fagen**
Director of Operations

**Cassandra Marketos**
Advisor, Community Engagement

**Tim Gavin**
Community Product Manager

**Katie Harbath**
Community Advisory Board Chair

**Samidh Chakrabarti**
Advisory Board

# Our Fellows

# We have validated key parts of this model

▷ There is huge need and demand for this expertise
  ➢ Governments, NGOs, advocacy groups, academics, and platforms
▷ Our members can meet this demand
  ➢ They have the necessary knowledge and experience
  ➢ However, their time is precious and limited
▷ We need more staff to unlock this impact
  ➢ **Add scale**: triage and prioritize the flood of interest
  ➢ **Add structure:** maintain and execute clear async workflows
  ➢ **Add speed:** help push out published artifacts across all active workflows
  ➢ **Add strategy:** understand what are the most important problems to tackle
  ➢ **Add support:** nurture the community and its members

# How We Deliver Impact

▷ Social internet companies **do Integrity better**
  ➢ Through legislation, regulation, self-regulation, compliance, etc.
  ➢ We shine a clear light on the *what*; we provide guidance on the *how*

▷ People **understand Integrity better**
  ➢ We give shape to the idea of an **Integrity worker**
  ➢ We bring together various Integrity roles (e.g. trust and safety, anti-spam, anti-ad fraud, threat intelligence, feed ranking) **under a shared identity, working on a shared project**

▷ Integrity Professionals have **more power**
  ➢ Integrity workers, by and large, know *how* to fix the social Internet, but they often lack the power and position inside companies to do so
  ➢ We advocate for companies to change their organizational design to **give these professionals the means and the power to solve these problems** in an ongoing fashion
  ➢ The role of "doing the right thing" is **honored, not silenced**

# Our Recent Accomplishments



▷ We advised **"Platform Accountability and Transparency Act"** authors on how platforms should report harms that occur and make their algorithms more transparent and explainable

▷ We advised the signatories of the **"Code of Practice on Disinformation"** on the quantitative reporting to demand from platforms

▷ Our community rallied in defense of integrity professionals when they **came under fire by Elon Musk**

▷ We have built **tracking dashboards to hold platforms accountable** for the content they most amplify

# Funding & Growth

We have raised **over $1M** to date from foundations and small donors

We are committed to remaining **independent of grants from platforms**

**To implement the work already underway,** we need an additional **$1M in 2022, $2.8M in 2023, and $3.6 M in 2024**

**We have other time-sensitive, high-impact opportunities that can be unlocked** if we receive additional funding beyond our current commitments (est. $2M or more)