Neither Band-Aids nor Silver Bullets: How Bug Bounties Can Help the Discovery, Disclosure, and Redress of Algorithmic Harm

Camille François | @camillefrancois
Dr. Sasha Costanza-Chock | @schock

USENIX Enigma ‘22
WHO ARE WE?

CAMILLE FRANÇOIS
CRASH Project Co-Lead, Algorithmic Justice League
(Presenter & Co-Author)

DR. SASHA COSTANZA-CHOCK
Director of Research & Design, Algorithmic Justice League
(Presenter & Co-Author)

JOSH KENWAY
Bug Bounties Research Fellow, Algorithmic Justice League
(Co-Author)

DR. JOY BUOLAMWINI
Founder & Exec. Director, Algorithmic Justice League
(Co-Author)

INIOLUWA DEBORAH RAJI
AI Harms Research Fellow, Algorithmic Justice League
(Co-Author)
Read the report at:
ajl.org/bugs
Algorithmic Harms

When an actor, such as an individual or an institution, uses an algorithmic system to automate classification, prediction, recommendations, scoring, etc., and as part of a process that causes people harm, such as loss of opportunity, violation of rights, affronts to dignity, social stigma, loss of freedom, physical safety, or loss of life.

(AJL Working Definition)
In 2020, the ACLU announced they are suing on behalf of Robert Williams in Detroit, who was falsely arrested in part based on failure of facial recognition technology.
MEET OUR COLLABORATORS AND INTERVIEWEES

RAYNA STAMBOLIYSKA
Fmr. VP – Governance, YesWeHack

MARCIA HOFMANN
Digital Rights Attorney

ALEX RICE
CTO, HackerOne

AMIT ELAZARI BAR ON
Director, Global Policy
Intel Corporation

JACK CABLE
Security Researcher

RYAN ELLIS
Associate Professor, Northeastern University

MARTEN MICKOS
CEO, HackerOne

YUAN STEVENS
Tech Policy Research Lead, Ryerson University

LISA WISWELL-COE
Fmr. Project Manager, Hack the Pentagon

MATT GOERZEN
Researcher, Data + Society

DINO DAI ZOVI
Security Researcher

ALGORITHMIC JUSTICE LEAGUE | @AJLUnited
PRE-DIGITAL BUG BOUNTIES

BUG BOUNTIES GO MAINSTREAM

BOUNTIES FOR ALGORITHMIC HARMS

ALGORITHMIC JUSTICE LEAGUE | @AIUnited
## Design Levers Across the Bug Bounty Spectrum

<table>
<thead>
<tr>
<th>Reporting Model</th>
<th>Voluntary</th>
<th>Adversarial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compensation Model</td>
<td>Non-Money</td>
<td>Bounties</td>
</tr>
<tr>
<td>Disclosure Model</td>
<td>Full Disclosure (incl. Delayed Full Disclosure)</td>
<td>Coordinated Disclosure</td>
</tr>
<tr>
<td>Participation Model</td>
<td>Public</td>
<td>Private / Invite-Only</td>
</tr>
<tr>
<td>Program Management</td>
<td>Platform-Managed</td>
<td>Mixed Management</td>
</tr>
<tr>
<td>Program Duration</td>
<td>Ongoing</td>
<td>Time-Limited</td>
</tr>
<tr>
<td>Program Scope &amp; Access</td>
<td>Constrained ↔ Expansive</td>
<td>‘Closed Box’ ↔ ‘Open Box’</td>
</tr>
</tbody>
</table>

**Algorithmic Justice League** | @AJLUnited
The Artist who can make an instrument that will pick or open this lock shall receive 200 guineas the moment it is produced.

Applications in writing only

Bramah’s PATENT LOCKS

The Public respectfully informed that every lock made by BRAMAH & Co. will be stamped with their address, 124, Piccadilly

Image Source:
Cabinet Magazine / Science and Society Picture Library
# BRAMAH’S CHALLENGE LOCK

<table>
<thead>
<tr>
<th>Reporting Model</th>
<th>Voluntary</th>
<th>Adversarial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compensation Model</td>
<td>Non-Money</td>
<td>Bounties</td>
</tr>
<tr>
<td>Disclosure Model</td>
<td>Full Disclosure (incl. Delayed Full Disclosure)</td>
<td>Coordinated Disclosure</td>
</tr>
<tr>
<td>Participation Model</td>
<td>Public</td>
<td></td>
</tr>
<tr>
<td>Program Duration</td>
<td>Ongoing</td>
<td></td>
</tr>
<tr>
<td>Program Scope &amp; Access</td>
<td>Constrained</td>
<td>Expansive</td>
</tr>
</tbody>
</table>
Wearing Many Hats

The Rise of the Professional Security Hacker

Matt Goerzen
Gabriella Coleman

Bounty Everything

Hackers and the Making of the Global Bug Marketplace

Ryan Ellis & Yuval Steiner
Trust in Facebook has dropped by 66 percent since the Cambridge Analytica scandal
Sixty-five percent of survey respondents say they want Facebook to disclose how it uses the personal information it collects.

After data incidents, Instagram expands its bug bounty

Instagram ad partner secretly sucked up and tracked millions of users’ locations and stories

Google to pay security researchers who find Android apps and Chrome extensions misusing user data
Rockstar Games
Publishers of such popular games as Grand Theft Auto, Max Payne, Red Dead Redemption, L.A. Noire, Bully & more.

Targeted Opportunity: GTA Online and Red Dead Online - Incorrect Ban Bounty

We occasionally receive support tickets from people who have been suspended for cheating in Grand Theft Auto Online or Red Dead Online, denying that any cheating took place and claiming a “false-positive”. We take these claims seriously and investigate them thoroughly. However, we have yet to find any evidence to support these claims.

We are offering a $10,000 bounty for any researcher who can successfully identify a reproducible incorrect ban in GTA Online or Red Dead Online. The types of activities that people have claimed (but we do not believe) to have caused bans include: using legitimate visual overlays (e.g. FPS) or audio chat integrations, other players spamming the in-game reporting system and having cash/rank forced on them by modders, but please do not limit yourselves to investigating these methods.

Our focus is on continuing to ensure that our anti-cheat system does not ban anyone who is playing the game normally and consistently with our terms of service. Any reports that are inconsistent with the spirit of what we are focused on will be ineligible for a bounty and closed. Furthermore, no complaints or commentary about existing bans will be reviewed. Note that bans incurred as a result of violating our terms of service will not be reversed. However, if anybody were to successfully illustrate a reproducible inappropriate ban, any banned test accounts would of course be restored.

As stated above, in order to participate please submit our enrollment form and await an invitation to our private bounty program. We will not accept reports pertaining to an in-game bounty or incorrect ban bounty that are submitted to the public program page.
Before

After

ALGORITHMIC JUSTICE LEAGUE | @AJLUnited
### Twitter’s Bias Bounty

<table>
<thead>
<tr>
<th>Reporting Model</th>
<th>Voluntary</th>
<th>Adversarial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compensation Model</td>
<td>Non-Monetary</td>
<td>Bounties</td>
</tr>
<tr>
<td>Disclosure Model</td>
<td>Full Disclosure (incl. Delayed Full Disclosure)</td>
<td>Coordinated Disclosure</td>
</tr>
<tr>
<td>Participation Model</td>
<td>Public</td>
<td>Private / Invite-Only</td>
</tr>
<tr>
<td>Program Management</td>
<td>Platform-Managed</td>
<td>Mixed Management</td>
</tr>
<tr>
<td>Program Duration</td>
<td>Ongoing</td>
<td>Time-Limited</td>
</tr>
<tr>
<td>Program Scope &amp; Access</td>
<td>Constrained $\leftrightarrow$ Expansive</td>
<td>‘Closed Box’ $\leftrightarrow$ ‘Open Box’</td>
</tr>
</tbody>
</table>
How teachers are sacrificing student privacy to stop cheating

Surveillance software for remote learning is drawing criticism from privacy advocates and lawmakers.

By Sara Morrison and Rebecca Heilweil | Dec 18, 2020, 9:30am EST

Proctor Ninja

Wave Rake: Proctorio

December 15, 2020

A zero-knowledge encryption flaw, enabling trivial brute forces of student data when Canvas/Moodle is used.

Students of color are getting flagged to their teachers because testing software can’t see them

Proctorio reportedly uses facial detection software that failed to recognize black faces over half the time

By Mitchell Clark | Apr 8, 2021, 8:34pm EDT

University will stop using controversial remote-testing software following student outcry

It's discontinuing the service after this summer

By Monica Chin | @mosquared98 | Updated Jan 29, 2021, 10:27am EST

Proctorio

Proctorio's facial recognition is racist.

March 18, 2021

Also, it kind of sucks in general.

Proctorio Is Doubling Down On Lawsuits Against Its Critics

The exam surveillance company is taking its legal crusade against critics a step further after coming under fire from students and privacy activists.

By Todd Feathers

ALGORITHMIC JUSTICE LEAGUE | @AILUnited
# Adversarial AI Bias and Harm Bounty

<table>
<thead>
<tr>
<th>Reporting Model</th>
<th>Voluntary</th>
<th>Adversarial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Compensation Model</strong></td>
<td>Non-Monetary</td>
<td>Bounties</td>
</tr>
<tr>
<td>Disclosure Model</td>
<td>Full Disclosure (incl. Delayed Full Disclosure)</td>
<td>Coordinated Disclosure</td>
</tr>
<tr>
<td>Participation Model</td>
<td>Public</td>
<td>Private / Invite-Only</td>
</tr>
<tr>
<td>Program Management</td>
<td>(Third Party) Platform-Managed</td>
<td>Mixed Management</td>
</tr>
<tr>
<td>Program Duration</td>
<td>Ongoing</td>
<td>Time-Limited</td>
</tr>
<tr>
<td>Program Scope &amp; Access</td>
<td>Constrained ↔ Expansive</td>
<td>‘Closed Box’ ↔ ‘Open Box’</td>
</tr>
</tbody>
</table>

*Algorithmic Justice League | @AILUnited*
Thank you!

To learn more:
https://ajl.org/bugs