



# Assume Worst Intent

PyCon UK 2018  
Twitter: @alexwlchan  
Slides: alexwlchan.net

# Content warnings



- Online harassment & abuse
- Mentions of racism, misogyny, sexism, suicide, rape & death threats
- Brief mentions of other horrible things on the Internet

Saturday

\$100 request  
for Tahoe



Monday

\$100 for Tahoe



Yesterday

\$25 for Tickets



3:00 PM

\$100 for Tahoe



REQUEST

PAY



**Anna Marie Clifton**

@TweetAnnaMarie

A friend's abusive ex has been sending her \$1 on  
@SquareCash regularly for months... cause he can add a  
message & she can't block that.

11:31 pm - 22 Oct 2016

<https://twitter.com/TweetAnnaMarie/status/789957313649967104>

User safety can't be  
an afterthought.

Most developers  
mean well.

They don't build services  
to enable harassment.

If you allow user-to-user  
interactions, you have  
the possibility of  
harassment.

# What does online harassment look like?

- Sending nasty messages
  - spam, phishing, personal attacks, hate speech, threatening/inappropriate imagery, sexual/pornographic content, rape threats, death threats
- Posting personal info
  - outing, doxing, SWATing
- Sharing intimate photos without permission
- Grooming young people
- Identity theft
- Exclusionary behaviour
- Dog piling
- So many more

# **Bridge fall girl, 15, felt ‘pressure’ from social networking websites**

~ The Telegraph

# What does online harassment look like?

- Sending nasty messages
  - spam, phishing, personal attacks, hate speech, threatening/inappropriate imagery, sexual/pornographic content, rape threats, death threats
- Posting personal info
  - outing, doxing, SWATing
- Sharing intimate photos without permission
- Grooming young people
- Identity theft
- Exclusionary behaviour
- Dog piling
- So many more

Who's doing the  
harassment?



8	Jan 19	20:30	hash.db
33	Jan 17	20:30	hash.db
34272	Jan 1	10:39	Install.io
55	Jan 1	10:40	Install.sy
115	Jan 23	22:04	!escan.sh
78	JAN 19	22:27	!realhost

# Who is an at-risk user potentially worried about?

- Abusive partner or ex
- Family members
- Classmates
- Coworkers and ex-coworkers
- Friends and ex-friends
- That one-date weirdo
- Rogue sysadmin
- Oppressive regimes

People are more likely  
to be hurt by  
**people they know.**

This is all  
very upsetting.



It doesn't have  
to be this way.

# How can we protect our users?

## Some best practices

Making your service better  
for **vulnerable users** can  
make it better for  
**everyone.**

#1

Diversify the team

*\$ git log*

commit 56aa901

Author: Alex Chan <alex@alexwlchan.net>

Date: Tue Aug 21 21:32:23 2018 +0100

Rearrange some of the scripts in the repo

\$ git log

commit 56aa901

Author: Alex Chan <[alex@alexwlchan.net](mailto:alex@alexwlchan.net)>

Date: Tue Aug 21 21:32:23 2018 +0100

Rearrange some of the scripts in the repo

#1

Diversify the team

#2

Think carefully  
about name policies

# “Real name” policies

- Anonymity causes problems, right?

[6063 comments](#) [Reply](#)  
[Flat](#) | [Top-Level Comments Only](#)

### **Skills For Someone Else**

Date: 2016-03-18 03:18 am (UTC)  
(Anonymous)

What skills have you developed that you never use for yourself?

### **Re: Skills For Someone Else**

Date: 2016-03-18 06:32 pm (UTC)  
(Anonymous)

Making tea and coffee. I don't drink either, but culturally I want to be able to offer them to guests.

### **Re: Skills For Someone Else**

Date: 2016-03-18 09:15 pm (UTC)  
(Anonymous)

How do you learn to make coffee if you don't drink it?

Nonny who doesn't like the taste of coffee would like to know.

### **Re: Skills For Someone Else**

Date: 2016-03-18 09:18 pm (UTC)  
(Anonymous)

I guess by asking whoever you make coffee for? Which might sound stupid, but some people prefer strong coffee, others prefer weak.

# “Real name” policies

- Anonymous discussion is possible – but expensive
- People don't care about the reputation of their “real” name
- Defining a “real” name is tricky
- Connecting wallet names to identities can be its own source of harassment
- Pseudonyms are a good middle ground

#2

Think carefully  
about name policies

#3

Implement robust  
privacy controls

# Robust privacy controls

- Have a way to ban/block malicious users (individual and platform level)
- Blocking can trigger retaliation – so consider shadow blocking or muting
- Offer granular access controls
- Making it easy to change makes everybody more comfortable and likely to interact

[Home](#)[Moments](#)[Notifications](#)[Messages](#)

Search Twitter



Tweet



## Alex Chan

@alexwlchan

You are blocked from following @alexwlchan and viewing @alexwlchan's Tweets. [Learn more](#)

### Who to follow

[Refresh](#) · [View all](#)**Helen ST** @helenst[Follow](#)**George Hickman** @ghick...[Follow](#)**John Chandler** @john\_ch...[Follow](#)

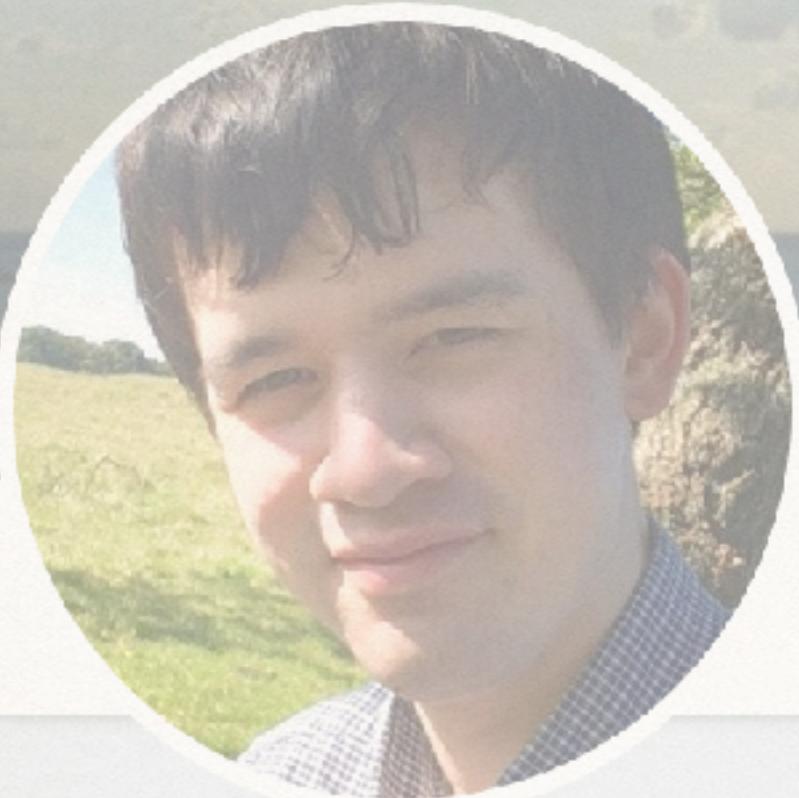
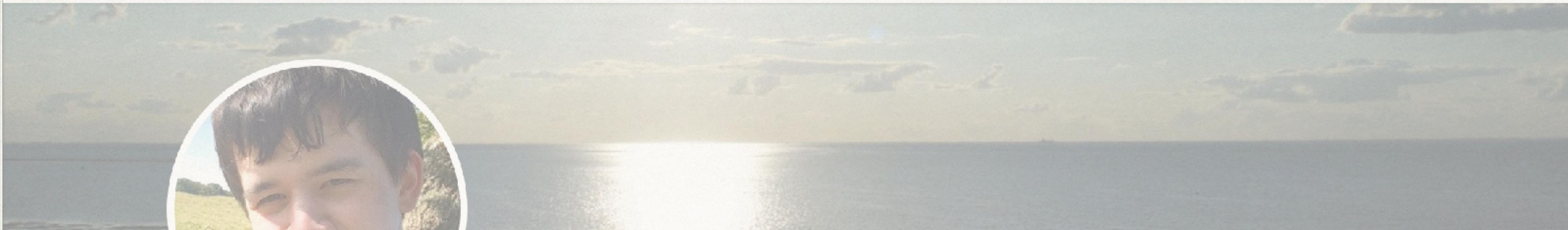
Find people you know

[Home](#)[Moments](#)[Notifications](#)[Messages](#)

Search Twitter



Tweet

**Alex Chan**

@alexwlchan

You are blocked from following @alexwlchan and viewing @alexwlchan's Tweets. [Learn more](#)

...

# You are blocked from following @alexwlchan and viewing @alexwlchan's Tweets.



George Hickman @gnick...

[Follow](#)

John Chandler @john\_ch... x

[Follow](#)

Find people you know

# Robust privacy controls

- Have a way to ban/block malicious users (individual and platform level)
- Blocking can trigger retaliation – so consider shadow blocking or muting
- Offer granular access controls
- Making it easy to change makes everybody more comfortable and likely to interact

## Privacy Settings and Tools

<b>Your activity</b>	Who can see your future posts?	Friends	<a href="#">Edit</a>
	Review all your posts and things you're tagged in		<a href="#">Use Activity Log</a>
	Limit the audience for posts you've shared with friends of friends or Public?		<a href="#">Limit Past Posts</a>
<b>How people can find and contact you</b>	Who can send you friend requests?	Everyone	<a href="#">Edit</a>
	Who can see your friends list?	Public	<a href="#">Edit</a>
	Who can look you up using the email address you provided?	Everyone	<a href="#">Edit</a>
	Who can look you up using the phone number you provided?	Everyone	<a href="#">Edit</a>
	Do you want search engines outside of Facebook to link to your Profile?	No	<a href="#">Edit</a>

<https://www.facebook.com/settings?tab=privacy>

# Robust privacy controls

- Have a way to ban/block malicious users (individual and platform level)
- Blocking can trigger retaliation – so consider shadow blocking or muting
- Offer granular access controls
- Making it easy to change makes everybody more comfortable and likely to interact

#3

Implement robust  
privacy controls

#4

Don't rely on  
technology to solve  
human problems

# Human moderation is best

- You don't know everything that happens (on your service or off)
- Context is important
- Give users a way to report problems, and believe them
- Look after your moderators



02920 180916



iMessage  
Today 10:42



Thinking of you!



iMessage



# Human moderation is best

- You don't know everything that happens (on your service or off)
- Context is important
- Give users a way to report problems, and believe them
- Look after your moderators

(This slide intentionally left blank)



## drum major for the drunken parade of life

@rahaeli

I screened volunteers for the LJ ToS team heavily, gave multiple layers of warning about what they'd see including actual examples during vetting, mandated breaks for people doing more than a few hours a week, and we still lost 40% of new volunteers within 3 months.

11:31 pm - 22 Oct 2016

<https://twitter.com/rahaeli/status/1036304125418504192>

# Human moderation is best

- You don't know everything that happens (on your service or off)
- Context is important
- Give users a way to report problems, and believe them
- Look after your moderators

#4

Don't rely on  
technology to solve  
human problems

#5

**Design with abusive  
personas in mind**

# How can we protect our users?

1. Diversify the team
2. Think carefully about name policies
3. Implement robust privacy controls
4. Don't rely on technology to solve human problems
5. Design with abusive personas in mind

You should always ask:

**How could this be  
used to hurt someone?**

**How could an  
abusive ex misuse this?**

# Assume Worst Intent

PyCon UK 2018

Twitter: @alexwlchan

Slides: alexwlchan.net

