(Anti) Social Media

How online services are exploited for harassment and abuse, and how to help

Alex Chan
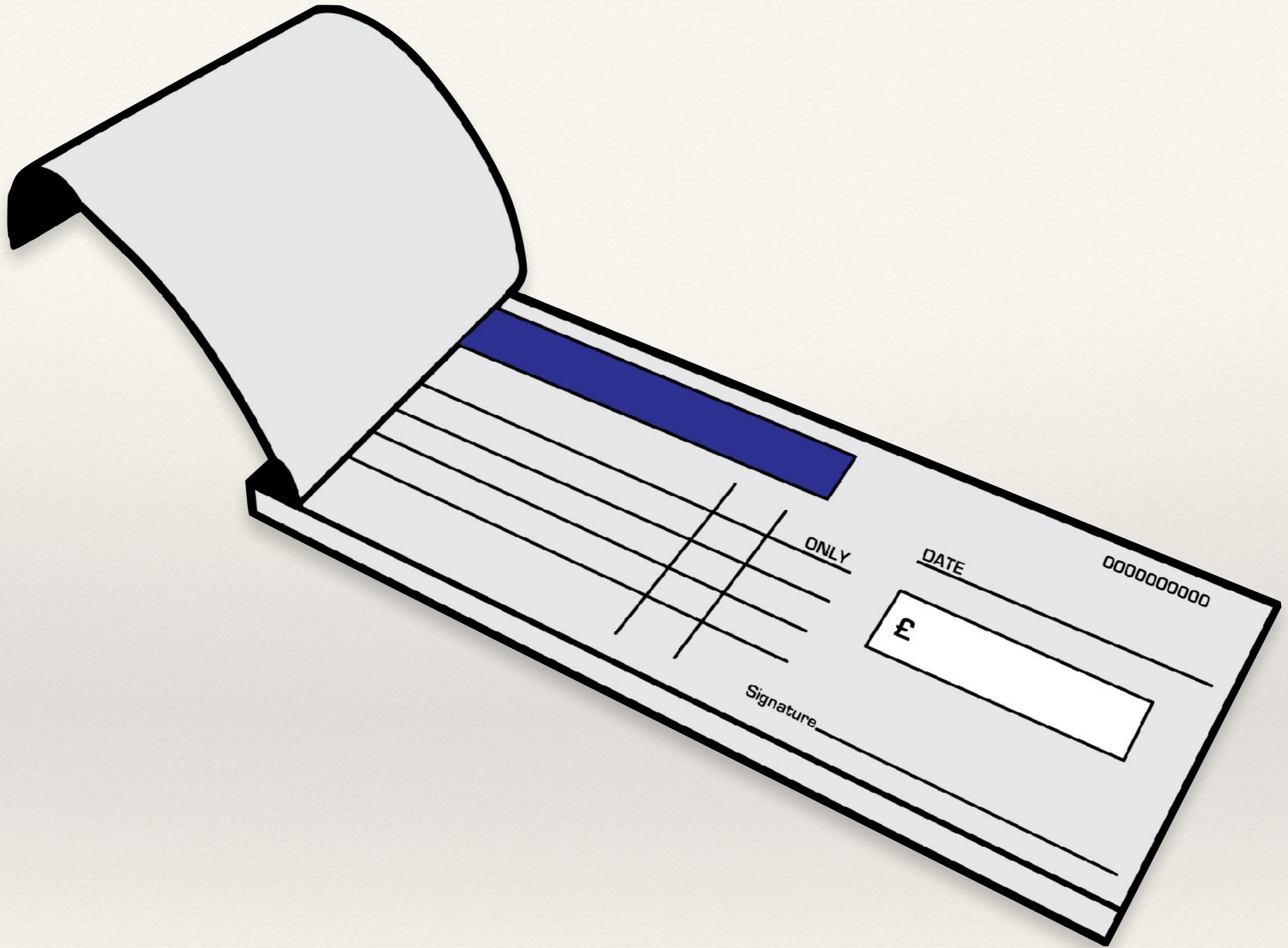
18 April 2018
Slides will be available after the talk

Content warnings ⚠️

- Online harassment

- Misogyny

- Racism

- Suicide

- Domestic abuse

- Police violence

- Sexual violence & assault

- Rape & death threats

ONLY

DATE

000000000

£

Signature

$100 request for Tahoe

$100 for Tahoe

$25 for Tickets

$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

Most developers **mean well**.

They don't build services to **enable harassment**.

Security protects **us** from **bad actors**.

Who protects **people** from **other people**?

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

What does online harassment look like?

Sending **nasty messages**
spam
phishing attacks

Virgin Media
@VirginMediaa1

5m ✓

Before we proceed, for security purposes of your account please provide the card number, expiry date, csc, card holder name, and billing address that's linked to the Virgin Media account. ^BP

5m

Sending **nasty messages**

spam

phishing attacks

personal attacks ("cyberbullying")

37%

UK children who said they "often" experienced cyberbullying

— Ditch the Label
Annual Bullying Survey 2016

Ryan Halligan loses his life to Taunts, Rumors and Cyber Bullying

Candlelight vigil held in honor of Megan Meier

Holly Grogan, 15, leapt to her death 'after abuse from Facebook bullies'

Sending **nasty messages**

spam

phishing attacks

personal attacks ("cyberbullying")

hate speech

threatening/inappropriate imagery

sexual/pornographic content

"I get emails that have my head photoshopped onto porn star's bodies, or dead animals, or brutally hurt women."

[Kelsey McKinney] gets two or three a week, usually under innocuous subject lines like "Read your piece" or "Just saying hello".

— Anna Merlan

Jezebel: "The Cops Don't Care About Violent Online Threats. What Do We Do Now?"

Sending **nasty messages**

spam

phishing attacks

personal attacks ("cyberbullying")

hate speech

threatening/inappropriate imagery

sexual/pornographic content

rape threats

death threats

Posting personal info

Posting **personal info** **outing**

Posting **personal info**

outing

doxing

Posting **personal info**

outing

doxing

SWATing

Destiny developer startled awake by police, sheriff's helicopter after faked 911 call

93 💬

By Brian Crecente | @crecenteb | Nov 7, 2014, 6:55am EST

Prank call sends close to 20 police officers to Southwest Portland home

Gamer who made "swatting" call over video game dispute now facing manslaughter charges

A Los Angeles man was charged with a Kansas man's death after making a false report to police in December.

By Emily Stewart | Jan 13, 2018, 4:58pm EST

Posting **personal info**

outing

doxing

SWATing

Sharing **intimate photos** without permissions (aka **"revenge porn"**)

Grooming young children

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**

This behaviour doesn't just affect online spaces

- Situations can escalate

- Situations that start online can move into physical space

- Even if it doesn't escalate, it's still unpleasant

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 is at risk?

Who is at risk?

- All your users

Who is at **increased** risk?

- Anybody in minority or an oppressed group
- Gender identity and expression
  Especially women, trans people

- Sexuality
- Religion or faction
- Race or colour

Who is at **increased** risk?

- Anybody in minority or an oppressed group

- Gender identity and expression
Especially women, trans people

- Sexuality

- Religion or faction

- Race or colour

- Anybody with controversial opinions
(e.g. "women and men should be paid equally")

"

[Diane Abbott] receives an incredibly disproportionate amount of abuse and was the target of almost a third of all abusive tweets we analysed.

— Amnesty International

Who is at **increased** risk?

- Anybody in minority or an oppressed group

- Gender identity and expression
Especially women, trans people

- Sexuality

- Religion or faction

- Race or colour

- Anybody with controversial opinions
(e.g. "women and men should be paid equally")

What are the main threat models?

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

- Oppressive regimes

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

78% / 57%

Number of female/male victims of domestic abuse who knew their attacker

— Home Office Statistical Bulletin

Homicides, Firearm Offences and Intimate Violence 2010/11

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

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

- Oppressive regimes

An abusive **current close contact**

- Parent, partner, sibling (very close)

- Coworker, friend (less close)

- Bad date, friend of a friend (new)

An abusive **current close contact**

- Physical access to you and your devices

- They probably know your passwords and security questions

- They know where you go online

- They can notice deviations from the norm

An **estranged close contact**

- Ex-partners (very close)

- Friend you've fallen out with, coworker from a previous job (close)

- Stalker

An **estranged close contact**

- No longer have physical access

- They still know you very well, including most of your "secrets"

- No revenge is too petty

A single or organised group of **bullies**

A disorganised mob and a **target of opportunity**

Angry **powerful person or group**

- What are your sysadmins up to? What about GCHQ?

Threat models

- Abusive current close contact

- Estranged close contact; stalker

- Single or organised group of bullies

- A disorganised mob and a target of opportunity

- Angry powerful person or group

This is all very upsetting.

It doesn't have to be this way.

How can we protect our users?

Best practices & advice

Why do high-risk users want to use a service?

- The same reasons as anyone else...

    fun
    work
    social life
    creative projects

- Plan an escape? Find support?

- Keep morale up?

Making your service better for **high-risk users** can make it better for **everyone**.

What are some possible approaches?

"If you're harassed online, **don't use the Internet**."

- It's 2018

- Many people need an **active** online presence

"I'm going to **ignore it** – if you don't like it, **use something else**."

- This is a morally dubious position

- What if they have friends on your service?

- Do their friends stick around?

Not everyone is **tech savvy**, and that's okay

- Not all **users** – people shouldn't need to be techies to be safe online

- Not all **attackers** – so some protection is better than nothing

How much data do you **need**?

- If you don't have the data, it can't be leaked or lost

- Why do you need my address, phone number, gender, favourite colour...?

Do you **need** to build that feature?

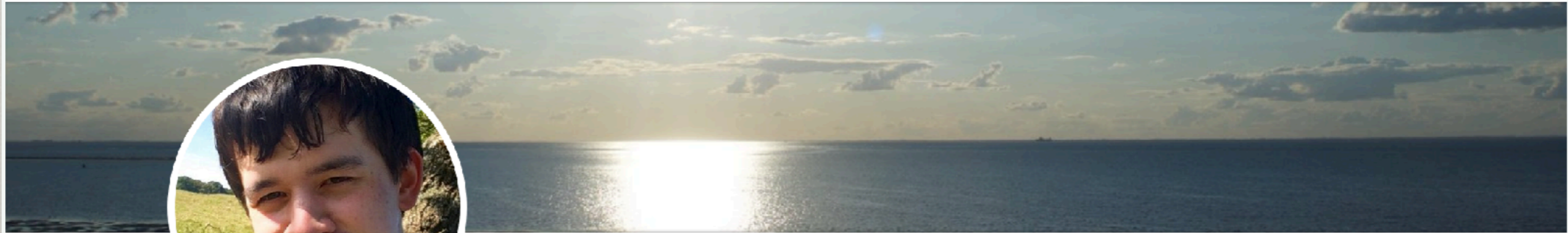- And if you do need it, how can you build it in a safe way?

Do you **need** to build that feature?

- And if you do need it, how can you build it in a safe way?

Have a way to **ban** or **block** malicious users

- Both at a platform and individual level
- **Every** platform should offer this

Alex Chan
@alexwlchan

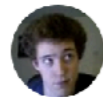You are blocked from following @alexwlchan and viewing @alexwlchan's Tweets. Learn more

Who to follow · Refresh · View all

Helen ST 🚴 💚 @helenst
Follow

George Hickman @ghick…
Follow

John Chandler @john_ch…
Follow

👥 Find people you know

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

Have a way to **ban** or **block** malicious users

- Both at a platform and individual level

- **Every** platform should offer this

- Sometimes the act of blocking is a trigger for retaliation

Allow "shadow blocking" dangerous users

- It may be safer to hide the fact that somebody has been blocked

Give your users **granular access controls**

- Fully public vs. fully private

- More granular controls are harder, but better

- Making it easy to change makes everybody more comfortable and likely to interact

Give your users **advanced security/privacy controls**

- e.g. Who can tag me in posts? Who can comment on my entries?

Privacy Settings and Tools

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

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

Give your users **advanced security/privacy controls**

- e.g. Who can tag me in posts? Who can comment on my entries?

- Not everyone uses them, but those who do **really** appreciate them

- People who need them will teach each other

Require **real names**?

- Anonymity is often a cause of problems

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.

Require real names?

- Anonymity is often a cause of problems
- It's possible to do truly anonymous discussion – but it's **hard**
- Defining a "real" name is tricky
- Connecting wallet names to identities can be its own source of harassment
- And problematic for people who change their names – married/divorced women, trans people

Require persistent names

- Pseudonyms are a good middle ground

Believe users when they say they have a problem

- Your lived experience is incomplete

- You don't know everything that happens on your service

- Don't punish somebody for reporting in good faith

Don't rely on **technology** to solve **human problems**

- Context is king, and you don't have it

iMessage
Today 10:42

Thinking of you!

iMessage

Don't rely on technology to solve human problems

- Context is king, and you don't have it

- Human moderators are expensive, but they're the **only** fix for human problems

- It helps to have moderators who speak the same language as your users

Design with abusive personas in mind

Good practices

- Not everyone is tech savvy – some protection is better than nothing

- Ask yourself: how much data do you need? Do you need to build that feature?

- Have a way to ban or block malicious users

  including shadow bans

- Give your users granular access controls

- Give your users advanced security/privacy controls

- Require persistent pseudonyms (maybe)

- Believe users when they have a problem

- Don't rely on technology to solve human problems

- **Design with abusive personas in mind**

This isn't easy.

You will have your own ideas.

You should always ask:

How could this be used to hurt someone?

How could an abusive ex use this?

(Anti) Social Media

contact info

@alexwlchan
alex@alexwlchan.net

slides & notes

alexwlchan.net

special thanks to

@azurelunatic
@ClareMacraeUK
@drakekin • @flexotron
@GailOllis • @HeofHIshirts
@jennyst • @jorm
@rahaeli • @tsunamino

for suggestions, inspiration & wisdom