

How to co-create content moderation policies: **The case of the AutSPACEs project**



Georgia Aitkenhead^{1†}, Susanna Fantoni[†], James Scott[†], Sophia Batchelor¹, Helen Duncan¹, David Llewellyn-Jones¹, Callum Mole¹, Otis Smith, Martin Stoffel¹, Robin Taylor, Kirstie Whitaker¹, **Bastian Greshake Tzovaras¹**

Overview




1. Background
2. Methods
3. Dilemmas
4. The AutSPACEs approach and Next Steps

Overview

1. Background



A photograph of a subway station platform. A train is blurred as it moves past the platform. People are standing on the platform, some looking towards the train. The station has a curved, vaulted ceiling and overhead lighting. A sign is visible above the platform.

**90% of autistic people process
sensory information differently
from non-autistic people**

Research to meet community priorities.

- **5.** Which **environments/supports** are most appropriate in terms of achieving the best life outcomes in autistic people?
- **9.** How can **sensory processing** in autism be better understood?



 Top ten questions for
autism research

 @AutSPACES

10.5281/zenodo.7550764



Research for real world Change.

AutSPACEs

Autism research into **S**ensory **P**rocessing for
Accessible **C**ommunity **E**nvironments.



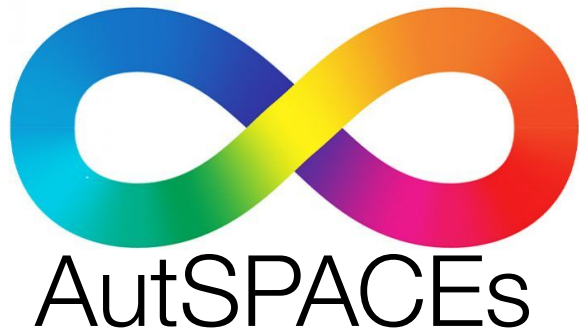
About AutSPACES

A research collaboration between

The Turing, Autistica, & Open Humans

The
Alan Turing
Institute





Collect qualitative data to improve our understanding of sensory processing in people's daily lives

Experience

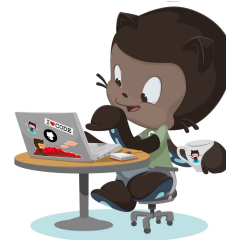
Working in the office can often be difficult. We have hot desking so I never know where I'm going to be sitting or who else will be there. The lights are very bright as well which can be difficult as it gives me headaches. It can also be loud which makes it difficult to concentrate on my work. The commute can also be stressful if the trains are crowded and I have to take the underground.

About AutSPACES

A **research collaboration** between
The Turing, Autistica, & Open Humans

A **community** of autistic people, their
supporters, researchers and open
source developers

The
Alan Turing
Institute



Open Life Science



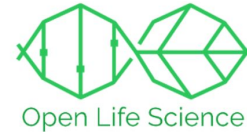
About AutSPACES

A **research collaboration** between
The Turing, Autistica, & Open Humans

A **community** of autistic people, their
supporters, researchers and open
source developers

A **citizen science** platform where
people can share their experiences of
sensory processing.

The
Alan Turing
Institute

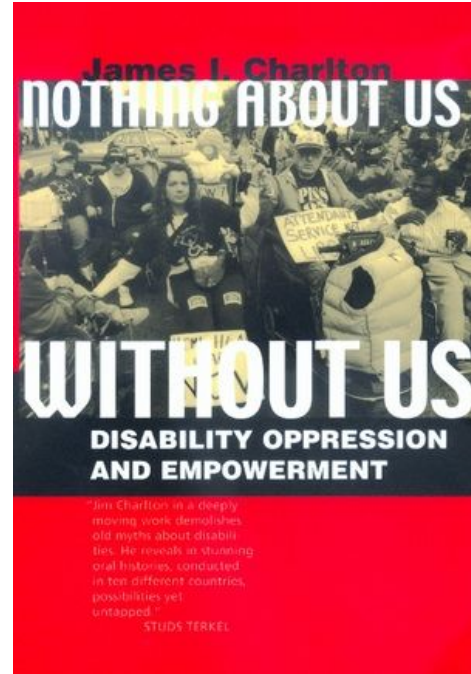


Open Life Science



Co-created citizen science

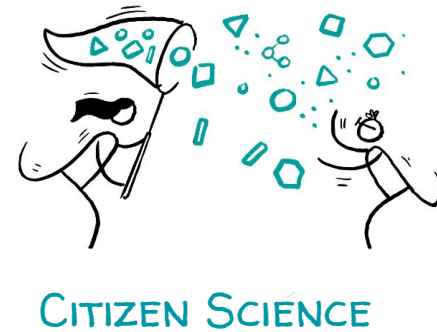
From co-design to co-ownership



Co-designing a Citizen Science Platform

Engaging people who are not professional scientists in scientific **research at scale**.

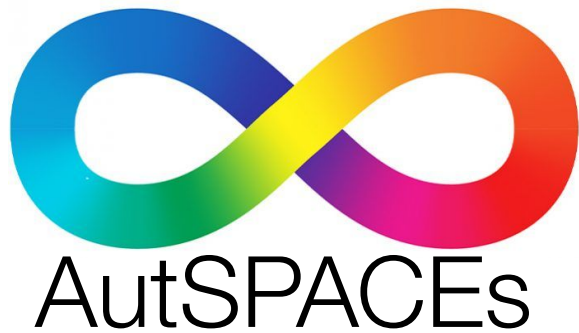
- Research design driven by a community.
- More reflective of the wider population.
- Real world experiences.
- Research results support the community.



Our co-design approaches

- **2 Scoping sessions** held with 26 autistic people, their supporters, and researchers.
- **3 Focus groups** held with autistic people, their supporters, researchers and developers.
- **1-2-1 co-working** with committed collaborators.
- Online **co-working sessions & community meet-ups**





Collect qualitative data to improve our understanding of sensory processing in people's daily lives



Share people's stories and adaptive techniques with others who have similar experiences

Educate neurotypical people to better support their friends, family and colleagues

Advise organisations on how they can design and adapt spaces to improve people's lives

Changed goals & scope

Share people's stories and **adaptive techniques** with others who have similar experiences

Educate neurotypical people to **better support** their friends, family and colleagues

Advise organisations on how they can design and **adapt** spaces to **improve** people's lives

Experience

Working in the office can often be difficult. We have hot desking so I never know where I'm going to be sitting or who else will be there. The lights are very bright as well which can be difficult as it gives me headaches. It can also be loud which makes it difficult to concentrate on my work. The commute can also be stressful if the trains are crowded and I have to take the underground.

Changed goals & scope

Share people's stories and **adaptive techniques** with others who have similar experiences

Educate neurotypical people to **better support** their friends, family and colleagues

Advise organisations on how they can design and **adapt** spaces to **improve** people's lives

Experience

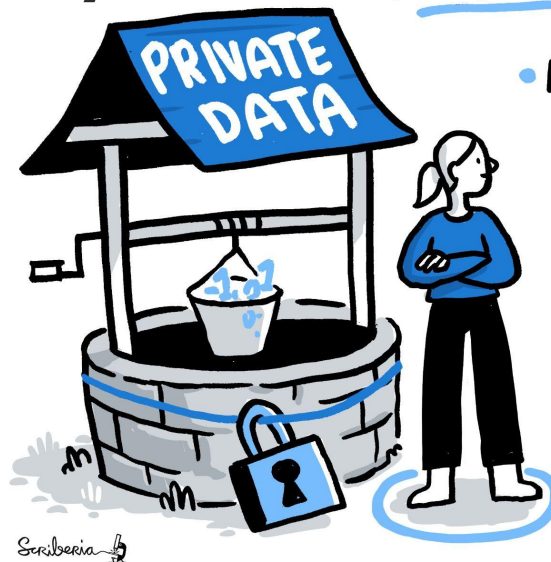
Working in the office can often be difficult. We have hot desking so I never know where I'm going to be sitting or who else will be there. The lights are very bright as well which can be difficult as it gives me headaches. It can also be loud which makes it difficult to concentrate on my work. The commute can also be stressful if the trains are crowded and I have to take the underground.

Recommendation

I try to get in early so that I can get my regular desk which is more in the corner away from the bright lights. I also often wear my headphones whilst working. I sometimes try and work non-standard hours to avoid the big crowds. And also walk instead of taking the underground.

Data Sharing & Agency

- Give people a way to share **some** data: dynamic **and** informed consent
- Moderate public experiences to keep everyone safe



PEOPLE DESERVE:

- DIGNITY
- AGENCY
- PRIVACY RIGHTS
- CONFIRMED CONSENT



@AutSPACES

10.5281/zenodo.7550764

Why Moderate?



Why Moderate?

“I think you have to work on the basis, unfortunately, that every single corner of the internet which doesn't have moderation just seems to fill up with Nazis, they're everywhere, and it happens in the most unlikely places, so I would say at least for the first couple of times you need to moderate the users”



Why Moderate?

- Marginalised groups more at risk of online abuse and to be victims of “false positive” moderation which targets them unfairly
- Social media use for autistic adolescents and adults is associated with better friendship quality
- Autistic people are likely to be at a higher risk of sexual abuse and exploitation online.
- Being ignored or left out of online interactions is associated with greater feelings of worthlessness and negativity
- Autistic people spend more time online than their non-autistic peers

(Feuston *et al.* 2020; Haimson *et al.* 2021; Salty 2021), (De Gregorio 2020)(Seering 2020), (van Schalkwyk *et al.* 2017), (Ward *et al.* 2018), (Landon 2016)Trundle *et al.* 2022), (Triantafyllopoulou *et al.* 2022),(Wang *et al.* 2020).



Why moderate?

can affect marginalised groups that might be at a particularly

- high risk of online abuse
- More likely to be the victims of “false-positives” where they are the target of moderation policies
- increasing calls for democratising moderation (De Gregorio 2020) and making use of community self-moderation (Seering 2020).
- On the side of benefits, studies have shown that the use of social media amongst autistic adolescents is associated with better friendship quality, an effect not seen in adolescents without autism (van Schalkwyk *et al.* 2017)
- autistic adults who use social media sites also report closer relationships than those who do not participate in online communication (Ward *et al.* 2018).
- autistic people are at higher risk of sexual exploitation (Landon 2016) as well as cyberbullying (Trundle *et al.* 2022).
- being ignored or left out of online interactions can lead to an increased feeling of worthlessness and negativity (Triantafyllopoulou *et al.* 2022).
- autistic people might be spending more time online than their non-autistic peers (Wang *et al.* 2020),

Overview

1. Background
2. Methods



Moderation Goals

A welcoming, inclusive, and safe online space for autistic people.

A space where autistic people can nevertheless share challenging experiences

Decisions which balance the complex needs of the community.

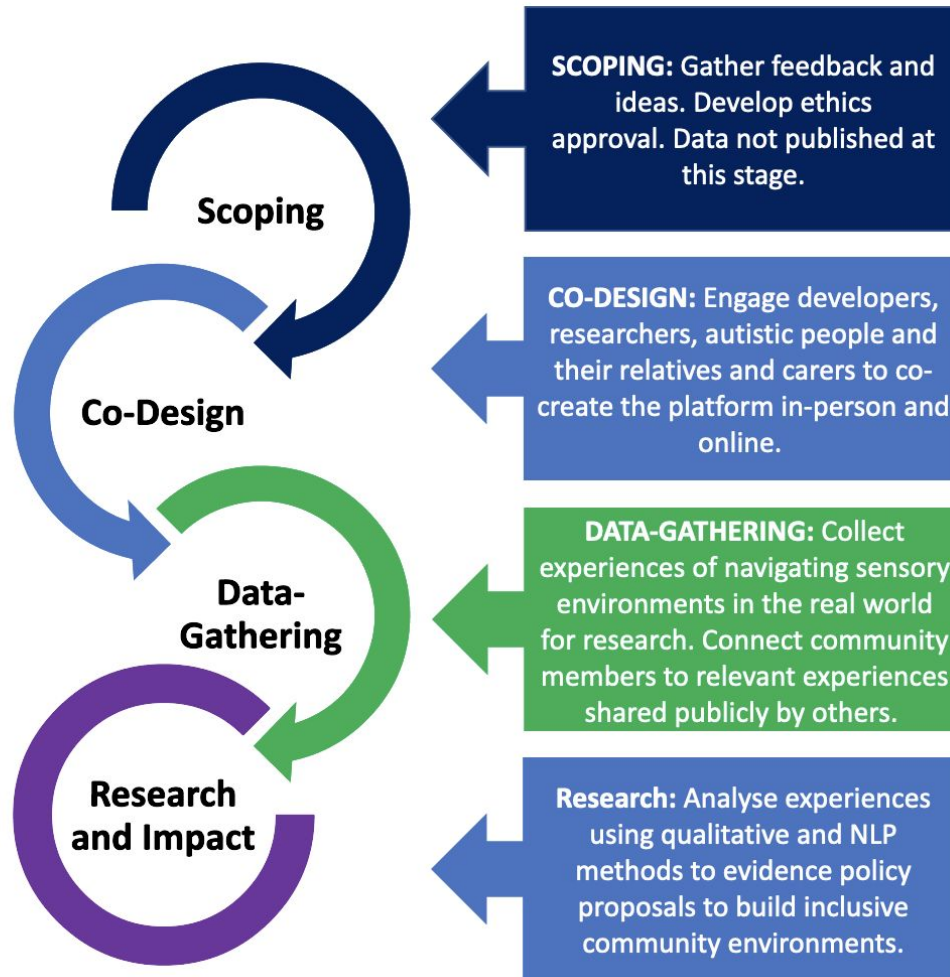


Building our inclusive online space

AUTSPACES
MODERATION AND
SAFETY



@AutSPACes
#MozFest



Moderation guidelines

To develop our moderation strategy we:

- Held focus groups
- Built user personas
- Reviewed and consolidated existing frameworks
- Engaged with the wider Autspace's community for feedback
- Reviewed academic literature and moderation best practice.



Overview

1. Background
2. Methods
3. Dilemmas



Major Challenges

- How can we meet users' freedom of expression whilst adhering to the project's ethos of a safe and inclusive environment?
- Finding balance between inclusivity and authenticity; to ensure an accurate research evidence base.



Commenting on others' experiences

"I think that will encourage good will, because it will encourage people to [give] feedback more."

"[it's] my experience, there's nothing to debate about it"

"Could cause arguments"



Triggering content

"...there may be people who are writing about abuse, or are writing about suicide attempts. Those are then triggers for others"

"...we have a duty of care"

"people sharing their own story is very powerful"

"awareness is really important – understanding"



Reporting on behalf of others

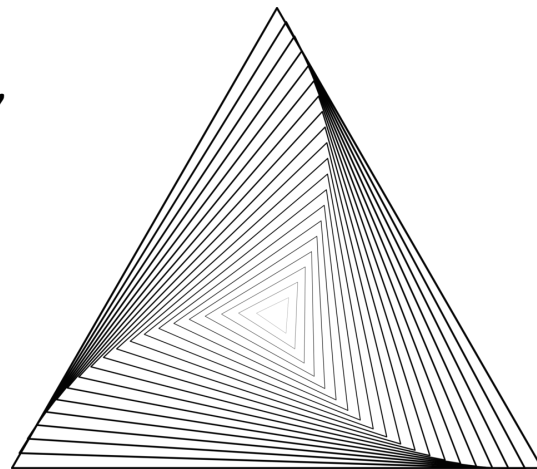
“[It’s] very difficult to make a place welcoming to autistic people when you also have a lot of neurotypical people explaining about autistic people”

“[My son’s] experiences would never be heard”



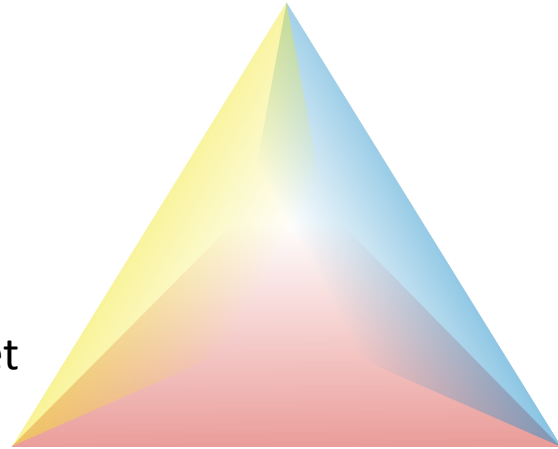
The Moderator's Trilemma

- 1) large & diverse userbases,
- 2) centralised & top-down moderation policies & practices
- 3) avoiding angering large parts of platform users



The AutSPACEs difference

- 1)** specific community of autistic users
- 2)** community co-designed and controlled moderation policies & practices
- 3)** balancing priorities to meet user needs



Overview

1. Background
2. Methods
3. Dilemmas
4. The AutSPACEs approach and Next Steps



Content submission and moderation guidelines with rules for content warnings

- 1) Transparency and equitability
- 2) Allows negative experience to be shared while minimising risk to vulnerable readers
- 3) Less burden on moderators to make subjective judgements
 - **co-created rules, uniformly applied**

Solutions & Innovations: Story Categorisation

Red, Green and Amber

- **Red** = Breaks CoC and will not be approved for public sharing
- **Green** = Meets all requirements
- **Amber** = Doesn't break CoC but contains potentially triggering content so will be flagged



We have excluded potentially triggering content, please select any categories that you would be comfortable seeing

- Allow all triggering content
- Abuse (physical, sexual, emotional and verbal)
- Drug and/or Alcohol misuse
- Negative body image
- Violence and Assault
- Mental Health Issues
- Other

Drug and/or Alcohol misuse

5. Self-medication

I once went to a party that was really loud. I was so uncomfortable there. Everyone was getting dru...

AutSPACes Contributor

Expand

Content submission and moderation guidelines with rules for content warnings

- 1) Transparency and equitability
- 2) Allows negative experience to be shared while minimising risk to vulnerable readers
- 3) Less burden on moderators to make subjective judgements
 - **co-created rules, uniformly applied**

No commenting on other users posts

- 1) Less pressure on users to seek social approval
- 2) Avoids prejudice and discrimination
- 3) Increases safety to allow sharing of vulnerable, personal experiences
 - **better user experience, better science.**

Guidelines for sharing stories about others

- 1) Emphasises autistic voices first
- 2) Addresses concerns about non-autistic misunderstanding and prejudice
- 3) Allows high needs autistic users to be heard
 - **more diversity, demands care and respect**

Parents and carers commenting on behalf of others.

- Can the person who parents and carers are writing about share their own story with support?
- Are you making any assumptions or inferences about the autistic person you are sharing on behalf of?
- Are you adding subjective and personal details which are your own?



It is better to write neutrally about the things you are observing. Focus on the experience of the autistic person you are writing on behalf of. Don't focus on your own experiences or the experiences of other people, except when it is relevant to the autistic person's experience. In particular, do not use this space for venting or complaining about autistic people.

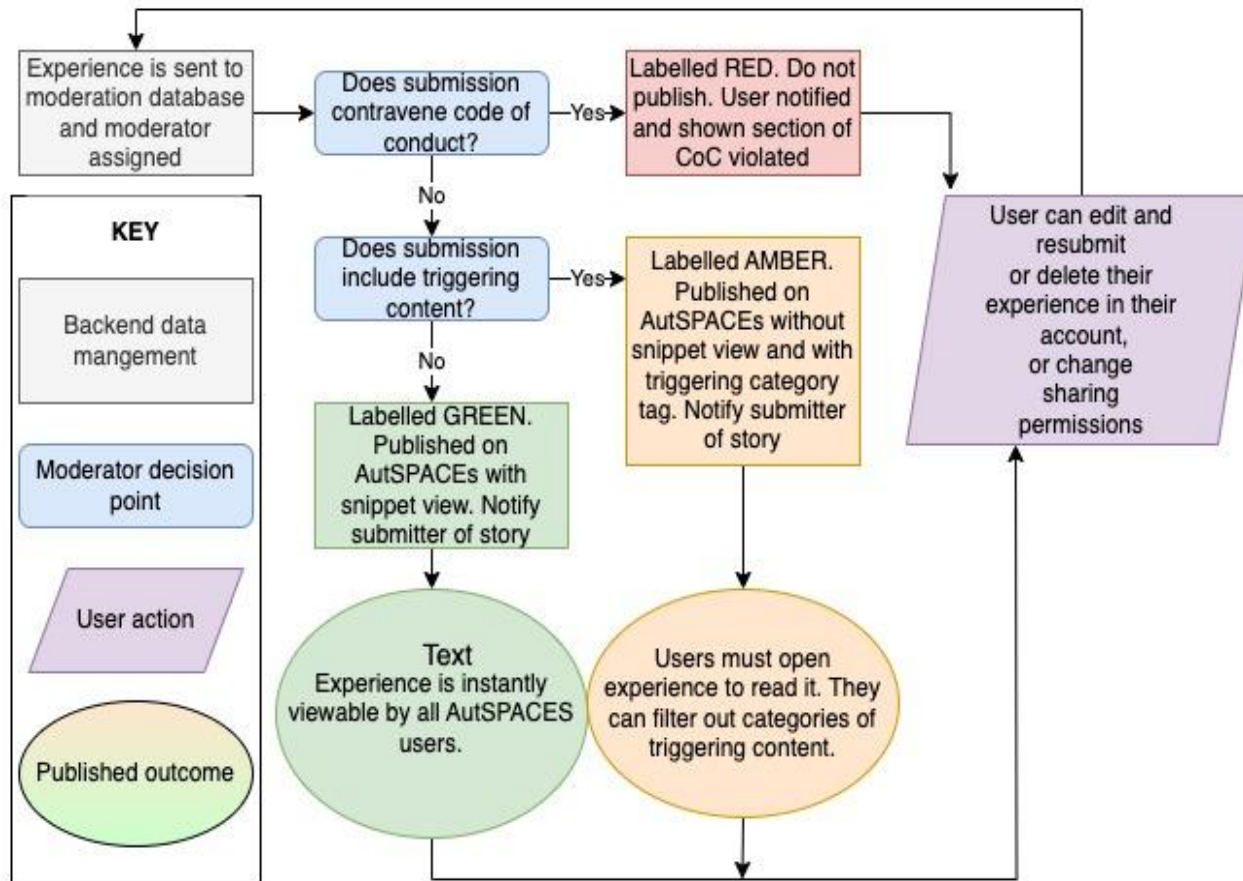
E.g. don't write, *"I hate it when my autistic son makes a scene in the supermarket, it's so embarrassing"*,

Please write, *"my autistic son cried and stood still in a busy supermarket. Some people around us showed negative expressions and body language while this happened. This usually occurs when we go to the supermarket."*



When to moderate: pre-publication review

- 1) Reduces risk for users
- 2) Gives moderators time to contemplate
- 3) Maintains AutSPACES' inclusive ethos
 - **less risk, higher standards**



Next Steps

- Implement moderation features on platform
- User test with a range of autistic people and carers
- Recruit autistic people as moderators
- Co-create training materials and resources
- Moderation processes will remain open and will be iterated once platform is live



Thanks to all the community members making these projects happen:

- KIRSTIE WHITAKER
- GEORGIA AITKENHEAD
- SUSANNA FANTONI
- BASTIAN GRESHAKE TZOVARAS
- JAMES SCOTT
- CALLUM MOLE
- HELEN DUNCAN
- JAMES CUSACK
- MARTIN STOFFEL
- DAVID LLEWELLYN-JONES
- JAMES KIM
- ROBIN TAYLOR
- SOWMYA RAJAN
- SOPHIA BATCHELOR

- MALVIKA SHARAN
- TOM AUGUR
- LOTTY COUPAT
- SOLEDAD LI
- ANOUSHKA RAMESH

...and all the other amazing collaborators/contributors



Find us at:

github.com/alan-turing-institute/AutSPACES

Read the preprint:

<https://osf.io/preprints/socarxiv/c2xe7>

AUTISTICA

The
Alan Turing
Institute





Questions and Discussion...