BEST PRACTICES FOR

GENDER-INCLUSIVE CONTENT MODERATION

*By **Alice Hunsberger**, **Vanity Brown**, and **Lily Galib***

Grindr

Fair and equitable content moderation is a cornerstone of building trust and adoption of any platform with user-generated content. Our society's understanding of gender and expectations for inclusivity are rapidly expanding.

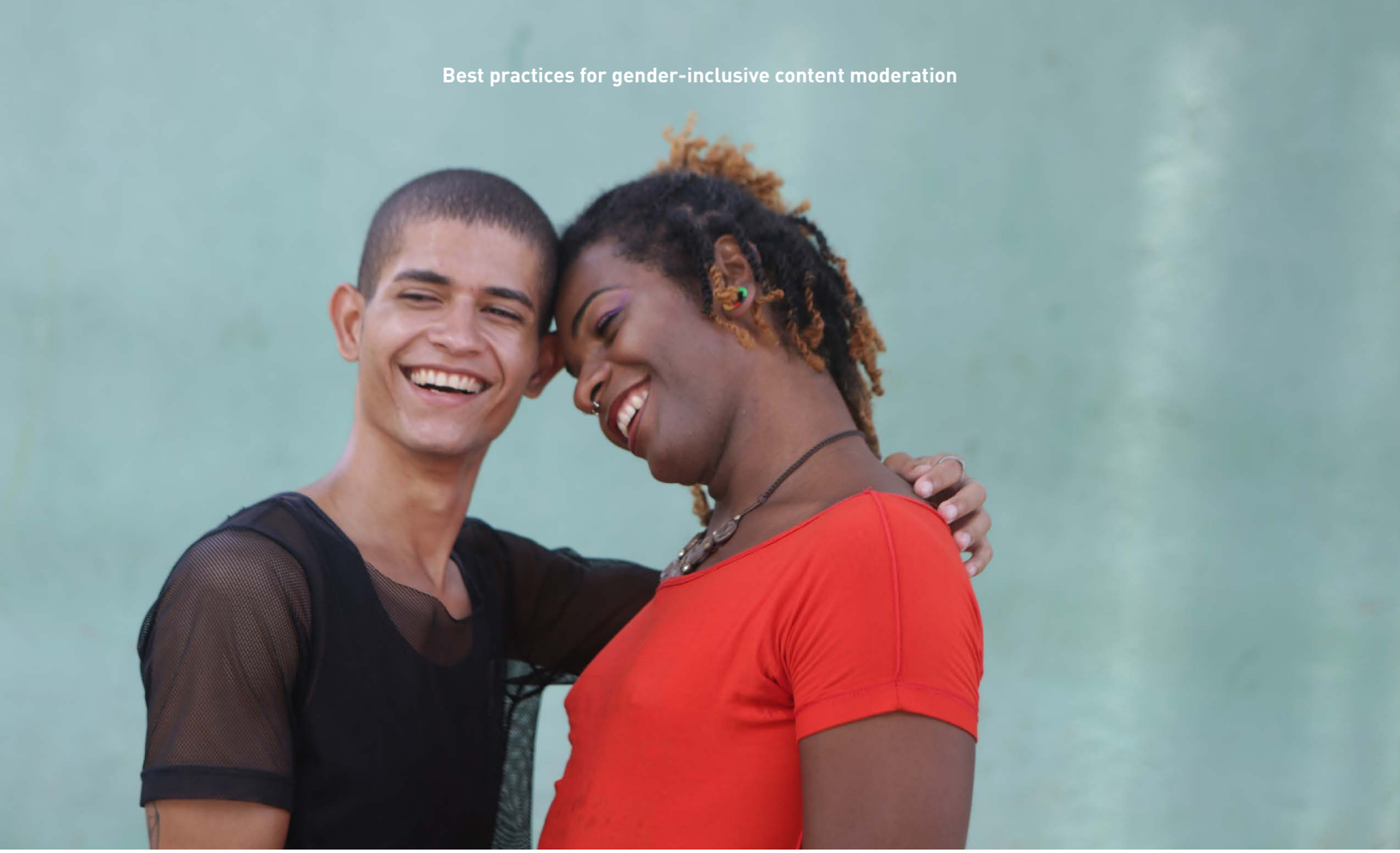
As **Trust and Safety** professionals in the social network and dating industry, we are honored to be a part of the evolution of content moderation strategies designed to have a meaningful impact on the lives of of trans, nonbinary, and gender nonconforming users[1].

It is not acceptable to shoehorn nonbinary people into binary rules and heteronormative standards.

Indeed, this population is growing and becoming more visible. Today, 1 in 6 Gen Z adults identifies as LGBT[2].

11% of the LGBTQ+ population is estimated to be nonbinary, making up 1.2 million nonbinary adults in the US alone[3]. As this population grows, it's increasingly important to make sure that your platform is welcoming, accessible, and moderated fairly.

[1] *For terms and definitions, see GLAAD's reference doc. Doing research on current terms is a key first step.*
[2] *From this Gallup survey*
[3] *UCLA Williams Institute 2021 study of Nonbinary LGBTQ adults in the US*



Grindr is the world's largest dating and social networking app for gay, bi, trans, and queer people and anyone who wants to connect with and be a part of the **Grindr** community. Naturally, we are focused on creating inclusive and forward-thinking moderation policies that honor the full expression of users' gender identity. That said, we strongly believe that having inclusive policies is necessary for all businesses – not just those that focus on LGBTQ+ people – so that all users can feel supported, included, and welcome.

In this whitepaper, we will outline strategies and best practices for creating thoughtful, equitable, and inclusive moderation policies and practices with gender inclusion in mind. This includes policy creation, product design, moderation, training resources for the moderation team, and user-facing resources. We'll offer insights about content moderation outside of the old-fashioned binary rules that have dominated our society and thus many platforms' practices.

We hope that this is a useful starting point, recognizing that this is not an exhaustive effort. Even small steps can build enough momentum to be meaningful to your users.

TABLE OF CONTENTS

DESIGN WITH SAFETY AND INCLUSION IN MIND

Careful product design sets your moderation team up for success by making sure that your product is inclusive and provides guardrails against common types of abuse.

- **Provide inclusive gender options.**

  Affirming gender is a critical part of creating parity on your platform for all of your users. Providing a gender field for male/ female/ other makes people who don't fit into the male/ female binary feel othered, excluded and less than. Even if you have a long list of gender identities to choose from, be thoughtful about how to leave some room open for those who don't identify with anything you provided. Consider using more open-ended options such as "ask me," "I'm not sure," and "a gender not listed here[4]" rather than "other."

- **Include pronouns.**

  The ability to list pronouns is important to trans and nonbinary people to prevent them from being misgendered or referred to incorrectly.

  Sixteen percent of LGBTQ+ people under the age of 35 use they/them pronouns[5]. Additionally, many allies also list pronouns to help foster inclusivity. While he/she/they are the most common pronouns, some people do prefer other options.

- **Use caution with open text fields.**

  Free text fields for pronouns and gender may seem like an easy solution, but open text pronoun fields can lead to jokey or even hateful entries that require active moderation to maintain an inclusive platform.

  We suggest reserving open text fields for areas of the product that allow users to provide unique, enriching content compared to pronouns where a fairly comprehensive list of inclusive options can be used.

[4] *See the National Transgender Discrimination Survey*
[5] *See 2021 LGBTQ+ Community Survey*

- **Design with moderation operations in mind.**

  User-generated content (photos, text, video, chat) can enrich your product and give users freedom to express themselves. Providing users with tools to conveniently flag, report, or block users who abuse your platform's policies can help set expectations from the beginning and establish your platform as inclusive and supportive for all people.

- **Allow flexibility.**

  Some people won't feel comfortable sharing their pronouns, sexual orientation, or gender publicly, as doing so can open them up to bullying or harassment. Therefore, those fields should be optional, or able to be hidden/shown to select people. It's also important to allow people to change their profile fields at any time. They may identify in different ways at different times.

  That said, if users change their gender, it may have implications for things like photo policies. See the case study on topless nudity policies (p.8) for examples.

Every platform must navigate the tension between a totally open system which requires a lot of moderation and a closed system which needs less moderation but is more restrictive. Further on in this guide we include information on creating effective moderation policy and resources for the moderation team.

BEST PRACTICES FOR INCLUSIVE MODERATION POLICIES

Having fair and equitable moderation on your platform first requires thoughtful rules to set the stage for what behavior is expected and allowed. Gender-based expectations are woven into the fabric of our society but are receiving fresh challenges in meaningful ways. Changing social norms can make content moderation especially difficult. Clear and transparent policies help set expectations with your users about what is acceptable and avoids misunderstandings about moderators' actions.

- **Publish Community Guidelines.**

  By articulating your company's values through your **Community Guidelines** and moderation policies and practices, you earn the user's trust in your platform. Given the disproportionate discrimination they face, we suggest including an expresss statement that trans and non-binary people are welcome on your platform.

  Without specifics, users may make assumptions about what is allowed or not allowed - for example, they may think that your platform is banning users just because they are trans, or they may think that transphobic hate speech is perfectly acceptable.

- **Explicitly disallow discrimination.**

  Your community guidelines should explicitly disallow discrimination and bigotry on your platform. It can also be helpful to explicitly say that users who falsely report other users based on their gender, orientation, or any other protected class could be banned themselves for misusing your moderation system.

- **Have gender-free image rules.**

  Moderation rules, especially for profile images, must be as fair and equitable as possible. Having separate rules for men, women, and nonbinary people is outdated and feels unfair and unequal.

  See our discussion below about the inequities of gender-based rules.

- **Be gender inclusive and specific.**

  If it's not realistic for your platform to have gender-free photo rules, make sure that your rules are fair and outline specific guidelines for nonbinary users. If you do have different rules for men and women, it's critical that trans men are treated as men and trans women are treated as women when enforcing these rules.

Also keep in mind that assessments of how sexual a photo is can be influenced by gender bias: how much skin someone shows, how much body hair they are expected to have, how the fat on their body is distributed, etc. Rules should be objective, free of bias, and as intuitive as possible.

Unfortunately, nonbinary people are often discriminated against and may assume that your policies or procedures are biased or discriminatory. Therefore, it is important to be particularly patient and earn users' trust.

- **Explain why.**

  Given the lingering sensitivities of an all-too-often marginalized community, it is important to articulate why certain moderation actions are taken. At **Grindr**, we include this statement:

  "The following is Allowed: People of all shapes, sizes, ethnicities, genders, and identities expressing their sexuality joyfully"

  in our user-facing photo rules to express what we DO want. And internally, we have a moderation vision statement outlining the "why" behind **Grindr**'s moderation policies around gender:

Grindr Believes that:

- All of our members deserve fair and equal treatment, regardless of gender, body type, or other physical traits. Discrimination has no place at **Grindr**.
- Moderation policy must be clear, easy to understand, and easy to enforce, without a lot of room for interpretation or bias.
- Gender cannot be reduced to a binary male/female - there are many people who are nonbinary, genderfluid, and transgender.
- We should not force users to disclose their gender publicly when they are not comfortable doing so.
- It is not our role to guess the gender of our users.

CASE STUDY: TOPLESS PHOTO POLICIES

Grindr continues to advocate for gender-inclusive and equitable moderation policies. We believe that for adult-oriented platforms, it is fundamentally unfair for topless profile photos to be moderated differently based on one's perceived or prescribed gender. Consider these six scenarios and the dilemmas that content moderators face:

1 A user uploads a photo of themselves topless in a swimming pool. In their profile, the user identifies as male.

- By almost all standards of social norms in the US, this is perfectly normal, and his photo is approved.

2 A user uploads a photo of themselves topless in a swimming pool. In their profile, the user identifies as nonbinary.

- How should a moderator decide whether to approve the photo if there are rules against female toplessness?
- Does the moderator base their decision on the perceived gender, regardless of how the user identifies? How do moderators make a determination based on their individual perception or assumption around gender?
- If the user does not identify as female, what should the moderator do if the user appears to have breasts? How do moderators account for the fact that different body types affect breast size (irrespective of gender)?

3 A user uploads a photo of themselves topless in a swimming pool. In their profile, when the photo was uploaded, the user profile did not specify their gender (wanting to be seen as they are without labels). The moderators assumed the user was male, and approved the photo. However, the user is a trans woman and later updates her profile that she identifies as female.

- Should the photo be rejected due to platform rules against female toplessness?
- Should all photos be re-moderated now that the user has changed their gender to apply a discriminatory, gender-based norm?
- At what point in a user's transition are photos no longer allowed to be topless?
- If a user identified as male prior to their transition and uploaded male-presenting photos, are these photos no longer allowed?

4 A user uploads a photo of themselves topless in a swimming pool. In their profile, the user identifies as genderqueer. Sometimes, the user identifies as male and presents as male. Sometimes the user identifies as female and presents as female. They often change their gender on their profile.

- Should the photo be re-moderated every time the user changes their gender to female on the platform?

5 A user uploads a photo of themselves topless in a swimming pool. In their profile, the user identifies as female. The user is in an area where it is legal and acceptable to sunbathe and visit public beaches topless, even as a woman.

- Should local laws and customs be taken into account when creating or applying these moderation rules? If so, how do you determine where such profiles may be acceptable and where they may not? The amount of work to keep an up-to-date list of local customs and laws is not feasible for small-to-mid sized companies. Also what if users from another jurisdiction have the capacity to see users in that region?

6 A user uploads a photo of themselves topless in a swimming pool. In their profile, the user identifies as male. The user is trans and has not had top surgery. The photo is of a man, not wearing a shirt, but with breast tissue.

- If moderators are making decisions quickly, will they read the gender of the user, or will they just reject the photo on sight?

In all these scenarios, the cis man gets quick and fair moderation (approval) of his photo, but no one else does. For many users, having a photo rejected is an emotional experience. They may feel personally rejected, judged, or discriminated against if their photo isn't approved. This is why **Grindr** is working across the industry to support adoption of rules and processes so that all people can enjoy **fair and equitable moderation** of their images.

FLAGGING AND REPORTING

Even with the most inclusive policies, community guidelines, and moderation procedures, people may still act in ways that are abusive or biased.

On some platforms, trans and non-binary users can be over-reported by other users due to bias and discrimination. Users may report a trans person because they hold limiting beliefs that trans people shouldn't be allowed on the platform. Or they may be offended that a trans person has sent them a like or a message, and will report an innocent interaction as "harassment." Trans people can also be reported for "impersonation" by people who consider their gender identity to be invalid.

Additionally, Artificial Intelligence and Machine Learning systems are often not designed with nonbinary and trans people in mind (among other historically marginalized groups). For example, many popular systems that classify images using AI don't have categories for nonbinary people, and don't consider nonbinary people when detecting nudity - they are only designed for "male nudity" and "female nudity." This risks legitimate photos of nonbinary people being incorrectly rejected as "sexual" or "nude".

Therefore, we offer these suggestions to support your critical work to safeguard systems from reinforcing bias and discrimination.

- **Don't auto-deny/ auto-ban.**

  As detailed above, neither AI systems nor user reports are completely accurate or free of bias. Automated systems can be best used to quickly approve what is allowed. This frees your human moderation team to focus on the nuanced decisions about what should be banned or removed.

  Likewise, automatic bans or warnings based solely on flags from other users may have the unintended consequence of perpetuating the reporting users' prejudice. Therefore, it is important to incorporate some level of manual review.

- **Consider the reporting flow.**

  Are you giving users context about why they might report a user? Users may mistake flagging an account as a legitimate way to signal "I'm not interested in this person/don't want to see them any more," which will disproportionately affect already marginalized groups. You may need to explicitly add an option for "I'm not interested" that doesn't send to the moderation queues.

- **Consider a separate moderation queue.**

  Because trans and nonbinary users are discriminated against frequently, they may end up with a lot of flags or reports from other users. The more flags a user has, the more likely they are to get banned by moderators, even if there are no violations or minor violations. It's easy for a moderator to see a lot of flags and assume a user has done something wrong, even if they haven't.

  For this reason, you may want to consider additional training or even a separate moderation queue for trans and nonbinary users, so that moderators can reduce bias and keep the context in mind when moderating. If accounts are automatically restricted or "shadowbanned" until flags are reviewed, prioritizing flags on trans and nonbinary users so they don't sit in a restricted status for too long helps prevent their accounts from being perpetually disabled by invalid or biased flags. Empower your moderation team with special training to recognize bias and avoid discrimination.

RESOURCES FOR THE MODERATION TEAM

Moderation and **Trust and Safety** teams are often seen as cost centers, but they are tasked with the important job of supporting user safety, upholding your company's values, and protecting your brand. This work is tough, moderators are expected to work quickly and accurately, and the content is often emotionally draining and sometimes can even be traumatizing.

As outlined in this whitepaper, it's critical to be proactive in supporting the frontline moderation team (and reducing the need for moderation in the first place) through thoughtful product design, inclusive policies, and resources. Additionally, a successful moderation team is one that is given the tools, training, and resources to do the best job possible.

- **Give moderators context.**

  Make sure that your moderation team sees the full context of user profiles when making decisions. A moderator is better able to detect discriminatory reports against a user if they can see the profile fields that make the user more vulnerable to discrimination.

  If you don't have gender-free image guidelines, display the user's gender fields for moderators reviewing images. Moderators cannot be given the impossible task of guessing or assuming gender when applying gender-based rules.

- **Train... and keep training.**

  Everyone has different life experiences which shape their perceptions. We cannot take for granted that every person is informed on topics of gender identity and sexuality. Creating training materials for your team will keep everyone on the same page. Make sure moderators understand nuances around current terms, phrases, and identities.

  Bias and discrimination training is incredibly helpful for moderators to see where they may need to make an extra effort. Trans competency training for moderators is most effective when it's part of a bigger, intersectional training program to combat other/all forms of bias. Trans people as a whole won't have an equitable experience on your platform if there are issues with other intersectional identities.

- **Create knowledge bases.**

  As part of the team training materials, develop and continually expand the team's knowledge-base on topics of diversity and inclusion, while providing resources like glossaries (including acceptable and unacceptable terms) to support consistency in the moderation process. Resources will need to be updated often, as new terms are created

and others may be dated and become less popular.

- **Calibrate.**

  It is difficult to write comprehensive rules for every scenario, so grey areas and edge cases will always require attention. Calibrations and team-wide discussions gets everyone on the same page, which creates more consistent moderation and a better customer experience. Additionally, calibration helps your team broaden their understanding and make inclusive moderation decisions.

- **Localize.**

  Language and terminology relating to gender and gender identity may be used in different ways across different regions and among different groups of people.

  Some terms, phrases, or ideas may need to be shown completely differently in the product based on the user's area. This does not mean that you should avoid talking about these topics in parts of the world where homophobia puts LGBTQ people in even more danger - in fact, it's even more important to provide an inclusive and safe space for these users.

  Moderation can be less effective if your moderation team is operating in a language or region which is unfamiliar. Cultural sensitivity training can help the moderation team understand and serve users in different regions. If your company is expanding to specific regions, talk to local users and employ local support and moderation staff as practicable. Regional support and advocacy groups can be invaluable to build internal awareness of specific issues.

CASE STUDY: LOCALIZATION

*From **Jack Harrison-Quintana**, **Director of Grindr for Equality**.*

For users of our platforms around the world, it's important to make gender fields inclusive of culturally-specific identities. For India, one identity that is important to include is **Hijra**. For many people who are unfamiliar with the cultural and history of the country, some **Hijras** may appear simply to be trans women. In fact, though, these two identities are not the same. It's important to know that not all male-assigned-at-birth Indians who express themselves in a more feminine way today are necessarily **Hijras** or trans women, and both of these terms should be respected.

The term **Hijra** is a generally accepted term, but choosing terms to be included in some regions may be even more complex because terms are changing, just as they are in the English-language world. Terms that are commonly used today may slide into being derogatory in the future.

Indonesia provides a good example of an identity term whose meaning has changed over time. Waria used to be a commonly used term inside and outside the LGBTQ communities. These days, however, the word is considered offensive by many of the male-assigned-at-birth feminine Indonesian people it seeks to describe. On the other hand, like English words such as transvestite, there are plenty of queers who still use it as an important part of their personal individual identity.

All of this ambiguity means that companies wishing to create an open and welcoming environment must work with representatives of these communities to ensure that what is put in place reflects the most current terms and those that are best for their specific platform.

RESOURCES FOR USERS

Educating your user-base proactively can help stave off potential issues. Not everyone is familiar with the latest terms or expectations, and that's ok. However, you want to set clear guidelines for your community and help support those who may not understand right away.

Those who experience harassment or discrimination because of their identity want to know that your platform listens, supports them, and will take action.

- **Partner with trans organizations.**

  Work with trans organizations to help prepare user-facing resources about the trans community.

  This is helpful for three reasons: firstly, your financial support helps these organizations that are changing the world and blazing paths forward for the community. Secondly, you're benefiting from their expertise - no need for you to reinvent the wheel. Thirdly, it's never a great look to write about a community without consulting and including them, so partnering with an appropriate organization gets you started on the right foot from the beginning.

- **Utilize Help Pages.**

  User friendly help pages can quickly address your user's questions and concerns. It can be helpful to link to further reading/resources on topics that your users may want to educate themselves more on, such as pronoun use and gender identity.

  Including tool tips or links to help pages within your product can be incredibly helpful. For example, in the area where you ask users to list their pronouns, you can include a link to a help page explaining what pronouns are and why they're important.

- **Create Blogs.**

  Addressing topics in a blog post can help provide insight on where your company stands on an issue. Blogs are great for timely announcements like new features or policies, or statements of support for current news issues.

- **Set expectations.**

  This can be a simple "community pledge" overlay at registration welcoming users and expressing what you expect from them from the start. You could also include your community guidelines in a link during an onboarding email.

- **Use warnings, nudges, or prompts.**

  The best time to educate users is before they potentially cause harm. Machine learning models can detect harassment and hate speech at the moment it's sent, and can intercept the message with a prompt to rephrase or reconsider what is being said.

- **Support victims.**

  Make it easy for users to report harassment or discrimination, and explicitly include it as a viable option for reporting another user in your reporting flow. Consider providing resources or help pages to users after they report someone, and acknowledge when a report has been sent.

  Depending on your platform, consider transparency about what happens to an account after it has been reported. This provides closure to the victim and lets them know that you took them seriously.

  You also, of course, need to respond to reports in a timely manner and be available via customer support channels to explain your policies and respond to complaints.

TO CONCLUDE

Principles of inclusion and an equitable experience for all users should be woven into product design from the beginning. Proactive work can minimize the busy workloads of content moderation teams. Policy creation, content moderation work, ethical tech, and diversity and inclusion work are connected and this problem space should be approached collaboratively and cross-functionally.

The time to implement gender-inclusive policies and features is now.

Gender fluidity has become more and more widely acknowledged in our society, and trans, nonbinary, & gender nonconforming people (as well as allies) want to see companies take action, so that their identities are reflected in the products that they are using. This isn't just the right thing to do ethically, it can increase user engagement, mitigate risk, and protect your brand reputation, as well as keep your platform relevant.

ABOUT THE AUTHORS

Vanity Brown is a **Senior Trust and Safety** professional with a knack for investigations and keeping online communities safe. She has been committed to the work of shaping policy and developing moderation processes in the tech space for 10+ years, including multiple dating apps like eharmony and **Grindr**.

Lily Galib is a **Trust and Safety** leader who strives to create nuanced, human-oriented moderation experiences at scale. She's interested in the ways that technology can bring humans together across vast distances and make the world feel like a more intimately connected place. She enjoys identifying patterns, digging into data, and eating burritos. When she's not in meetings discussing the nuances of thong photos, you can find her doing outdoor things.

Alice Hunsberger has over a decade's experience leading **Customer Support**, **Customer Experience** and **Trust & Safety** teams at two of the world's most popular dating apps: **OkCupid** and **Grindr**. She specializes in creating a human-centered user experience through policy development, cross-functional collaboration, and sensitivity to a diverse and global user-base. She is currently the **Senior Director of Customer Experience** at **Grindr**. You can find her on LinkedIn here.

OCTOBER 2021