EDGELANDS

CONCEVOIR DES ESPACES NUMÉRIQUES SÛRS AVEC EMPATHIE : GUIDE PRATIQUE

LANDSLANDS
EDGEEDGE
LANDSLANDS

CONTEXTE

Ce document représente l'aboutissement de huit semaines de travail collaboratif mené par une équipe de chercheur·euses, de mentor·es et d'artistes, enrichi par les contributions d'intervenant·es invité·es, dans le cadre de la phase « Pop-Down and Beyond » de l'Institut Edgelands.

Le guide pratique suivant est le résultat final du projet : il vise à fournir des recommandations pour la conception d'espaces numériques sûrs, fondées sur l'analyse de sources primaires, des études de cas et des données issues d'enquêtes. Nous espérons que ce guide pourra être utilisé par des développeur·euses, des chercheur·euses, des artistes, ainsi que par les utilisateur·ices des espaces en ligne eux-mêmes.

Si vous souhaitez en savoir plus sur le contexte complexe de la sécurité numérique, les parties concernées, ainsi que les définitions de « espaces numériques sécurisés » et de « sécurité en ligne », vous pouvez accéder à la version étendue de ce guide pratique ici (en anglais uniquement).

LANDSLANDS
EDGEEDGE
LANDSLANDS

REMERCIEMENTS

Ce guide pratique a été développée par Pulkit Mogra, Tatiana Lysova, Lilian Olivia Otero, Catherine Keegan, Nina Martin, Mmabatho Oke, Jessica McClearn et Giovanna da Custódia, sous la direction de Nina Baranowska, Daniel Odongo et Laura García Vargas. Nous tenons à remercier toutes les personnes qui ont répondu à notre enquête et contribué à façonner les recommandations pour la création d'un espace numérique sécurisé. Nos remerciements vont également aux intervenant·e·s invité·e·s qui ont participé aux sessions d'ancrage du Sprint de Recherche et partagé leurs réflexions et leur inspiration. La conception graphique et la mise en page visuelle ont été réalisées par Flavia Lozano et Larissa Oliveira.

NOTE DE SYNTHÈSE

Ce guide synthétise les observations issues d'analyses de contenu, d'études de cas et d'une enquête qualitative afin de proposer des recommandations pour la conception d'espaces numériques sûrs qui placent l'empathie et les besoins spécifiques des communautés au cœur de leur approche, plutôt que des logiques extractives centrées sur les plateformes.

➔ LE PROBLÈME

Contrairement aux espaces physiques sûrs, les environnements numériques ne disposent pas de marqueurs clairs de sécurité. Trois acteurs principaux façonnent la sécurité en ligne: les plateformes, les gouvernements et les communautés, mais les approches actuelles demeurent insuffisantes.

Les **PLATEFORMES** conçoivent leurs services pour un « utilisateur moyen » universalisé (généralement cisgenre, hétérosexuel, blanc, issu des classes moyennes ou supérieures et du Nord global), rendant invisibles les vulnérabilités des populations marginalisées. Les incitations économiques privilégient les indicateurs d'engagement au détriment du bien-être des utilisateur·ices.

Les **RÉGULATIONS GOUVERNEMENTALES** se concentrent étroitement sur la prévention de « préjudices tangibles » (contenus pédopornographiques, terrorisme, fraude), en négligeant la dimension subjective et contextuelle du sentiment de sécurité en ligne.

Les **COMMUNAUTÉS** elles-mêmes jouent un rôle essentiel mais fragile dans la gouvernance de la sécurité via des codes de conduite et une modération souvent bénévole, tout en restant structurellement subordonnées aux architectures des plateformes.

➔ PRINCIPAUX RÉSULTATS DE LA RECHERCHE

Notre analyse a identifié quatre conditions fondamentales de la sécurité numérique.

Les **CONDITIONS RELATIONNELLES** incluent le respect des limites personnelles sans exiger de confrontation, une modération efficace conçue comme un travail relationnel (et non comme une simple application de règles), ainsi que la réduction du coût émotionnel de la participation dans des contextes où les utilisateur·ices doivent constamment se défendre.

Les **CONDITIONS CULTURELLES** englobent des normes sociales fondées sur la dignité et l'absence de jugement, une inclusivité linguistique entre langues et communautés, ainsi que des réseaux de pair·es offrant une infrastructure protectrice aux groupes vulnérables.

Les **CONDITIONS PROCÉDURALES** impliquent des règles claires et appliquées de manière consistante, l'autonomie des utilisateur·ices sur leur visibilité et le suivi de leurs données, ainsi que des structures de gouvernance légitimes, ancrées dans les expériences vécues plutôt que dans des décisions arbitraires d'entreprise.

Les **CONDITIONS INFRASTRUCTURELLES** portent sur la fiabilité technique et l'intégrité des plateformes, incluant des pratiques transparentes en matière de données, des mécanismes de signalement efficaces et des dispositifs de responsabilisation en cas de défaillance des systèmes.

Ces thématiques sont suivies de recommandations clés hiérarchisées, distinguant les éléments indispensables, souhaitables et les signaux d'alerte, afin de créer des espaces numériques plus sûrs.

➔ CONCLUSION

Ce guide se veut à la fois un outil pratique et un appel à repenser les espaces numériques comme des environnements gouvernés par les communautés qui les habitent, où la sécurité est co-construite et non imposée par des architectures d'entreprise optimisées pour l'engagement au détriment du bien-être.
Nous reconnaissons les limites de notre recherche, notamment la nécessité d'intégrer davantage de perspectives techniques, d'impliquer un éventail plus large de communautés et de traduire ce travail dans des langues autochtones. Enfin, nous appelons à approfondir les réflexions autour du sujet de la sécurité en ligne.

GUIDE POUR IDENTIFIER, CRÉER ET GÉRER DES ESPACES SÛRS EN LIGNE

Ce qui suit constitue un guide des espaces sûrs en ligne, basé sur notre recherche.

→ ÉLÉMENTS OBLIGATOIRES

POLITIQUE ET DOCUMENTATION

Une documentation claire et bien formulée joue un rôle central dans le sentiment de sécurité au sein des espaces numériques. La sécurité augmente lorsque les règles sont explicites, accessibles et visibles avant même de rejoindre un groupe. Cette transparence permet aux individus de décider en connaissance de cause si un espace numérique correspond à leurs valeurs et à leurs besoins, et donc s'ils souhaitent y participer.

Les règles doivent définir clairement les comportements acceptables, les normes de communication et les conséquences associées. Cela réduit l'incertitude et renforce la confiance dans le fait que les comportements préjudiciables seront découragés. Le langage utilisé doit être accessible à l'ensemble des membres potentiels de la communauté.

Les bonnes pratiques en matière de règles communautaires abordent explicitement et interdisent le harcèlement, l'intimidation, les discours haineux, les contenus extrémistes et les formes d'abus identitaires (par exemple l'usurpation d'identité ou l'utilisation de l'IA pour créer des images inappropriées d'une personne sans son consentement). Si les règles sont ambiguës, un espace de discussion, de révision et de reformulation doit être prévu. Remplacer un langage large ou abstrait,

comme « communication écologique » ou « positive », par des attentes explicites en matière de respect, de non-jugement et de transparence rend les normes plus compréhensibles et applicables.

Les lignes directrices communautaires doivent être cultivées avec la participation des communautés concernées et réévaluées de manière continue et itérative à partir des expériences concrètes des membres.

La gouvernance des données doit être clairement définie, en précisant comment les données personnelles sont collectées, utilisées, stockées et protégées. Des pratiques éthiques en matière de données, des principes de protection de la vie privée dès la conception et une communication transparente et accessible sur les mesures de protection sont essentielles au sentiment de sécurité. Si un outil numérique est multilingue, l'ensemble des éléments, y compris ceux relatifs à la protection des données, doivent être traduits dans toutes les langues concernées.

Enfin, les règles ne sont efficaces que si elles sont appliquées de manière cohérente et équitable à tou·tes, y compris aux modérateur·ices et aux administrateur·ices. Une application inégale ou perçue comme arbitraire érode rapidement la confiance et accroît les sentiments d'exclusion ou de vulnérabilité.

ÉLÉMENTS TECHNIQUES

Des outils de sécurité robustes et à jour sont des composantes essentielles de la sécurité numérique.

Ils doivent toutefois respecter la vie privée, être sensibles au contexte et éviter une logique punitive. Les filtres de contenu, la suppression automatisée de contenus perturbants ou illégaux et des mécanismes de signalement efficaces contribuent au sentiment de sécurité. Le chiffrement, l'anonymisation, la prévention des fuites de données et des mesures de sécurité actualisées, telles que l'authentification à deux facteurs, sont également cruciales.

Cependant, la modération automatisée peut être excessive, supprimant des contenus inoffensifs faute de contextualisation ou en raison de règles trop strictes. Cela peut limiter la participation, réduire au silence des expressions légitimes et affaiblir la confiance envers la plateforme. Elle doit donc fonctionner en complément d'une supervision humaine, avec la possibilité de contester une décision automatisée en fournissant des explications contextuelles.

Les protections techniques doivent aussi prévenir les abus communautaires, l'exploitation de données privées et l'infiltration hostile de communautés sensibles ou vulnérables. Les dispositifs de sécurité sont plus efficaces lorsqu'ils limitent ces risques de manière proactive plutôt que de réagir une fois le préjudice survenu.

Les utilisateur·ices doivent pouvoir contrôler leur propre confidentialité. Les plateformes doivent offrir des outils permettant de gérer la visibilité des profils, de divulguer sélectivement des informations

personnelles et de choisir entre des modes de participation publics ou privés.

Les politiques imposant l'usage du nom réel peuvent être préjudiciables lorsque l'anonymat et la confidentialité sont essentiels à la sécurité. Pour de nombreuses personnes, notamment issues de communautés marginalisées ou de contextes politiquement sensibles, l'anonymat constitue une mesure de protection indispensable.

Enfin, les plateformes doivent intégrer de manière proactive des ajustements et des options répondant aux besoins spécifiques de certain·es individus ou communautés, afin que l'adaptation devienne la norme et non un privilège à négocier.

ÉLÉMENTS COMMUNAUTAIRES

IEn l'absence de frontières physiques, les utilisateur·ices s'appuient sur des indices numériques informels mais puissants pour évaluer si un espace est sûr. Ils et elles peuvent repérer des marqueurs visuels, lisant en quelque sorte le « langage corporel » d'une plateforme. Les pronoms dans les biographies, les symboles inclusifs, les avertissements de contenu ou les codes de conduite épinglés en haut d'un fil constituent des signaux immédiats indiquant l'existence de limites claires et d'un espace intentionnellement cultivé.

Ces signaux visuels sont renforcés par des indices linguistiques qui façonnent l'atmosphère interpersonnelle. Les espaces sûrs privilégient un langage inclusif du « nous » plutôt qu'une

rhétorique antagoniste du « eux contre nous », et utilisent fréquemment des indicateurs de ton afin de réduire l'ambiguïté. L'inclusivité linguistique et la reconnaissance de la diversité sont des éléments clés des règles et codes de conduite, particulièrement importants pour les personnes neurodivergentes, pour lesquelles une communication explicite réduit l'anxiété et les malentendus.

Afin d'éviter les infiltrations non souhaitées, les communautés peuvent instaurer un processus de vérification de base pour les nouveaux membres, par exemple une étape d'intégration demandant de reconnaître les valeurs, l'objectif et les limites de l'espace avant une participation complète.

En définitive, le signal le plus déterminant reste comportemental. Les membres observent la manière dont la direction et la modération opèrent concrètement, évaluant la sécurité à partir de la rapidité, de la cohérence et de la transparence des réponses face aux violations. Lorsque les discours haineux persistent ou que l'application des règles paraît arbitraire, les règles écrites perdent leur crédibilité et le sentiment de sécurité se dissout rapidement.

En cas de conflit, les plateformes devraient encourager, dans un premier temps, la résolution par la bienveillance et le dialogue constructif, réaffirmant ainsi le rôle de chacun·e au sein de la communauté. Toutefois, des mesures disciplinaires doivent être appliquées rapidement lorsque la résolution n'est pas possible.

➔ CHECK-LIST RECOMMANDÉE

INDISPENSABLE / EXIGENCES MINIMALES

- ◯ Règles et lignes directrices communautaires claires, explicites et accessibles, avec intervention en cas de non-respect
- ◯ Politique solide de protection des données
- ◯ Fonction de capture d'écran restreinte dans les communautés privées
- ◯ Outils fiables de signalement des contenus préjudiciables
- ◯ Anonymat optionnel
- ◯ Supervision des modérateur·ices et administrateur·ices
- ◯ Modérateur·ices humain·es (qui ont accès à un soutien psychologique et à des ressources adaptées
- ◯ Langage inclusif
- ◯ Mesures réactives et proactives contre les discours ou activités nuisibles
- ◯ Pour les plateformes multilingues, traduction complète de tous les éléments, y compris les accords utilisateurs et codes de conduite
- ◯ Capacité interne dédiée en cybersécurité, en particulier pour les communautés à risque

SOUHAITABLE

- ◯ Processus de sélection lors de l'intégration
- ◯ Mesures de protection et de confidentialité intégrées à l'infrastructure de la plateforme
- ◯ Ressources éducatives sur les comportements et l'étiquette en ligne
- ◯ Co-conception participative intégrant les besoins des communautés
- ◯ Agents de soutien ou modérateur·ices issus de la communauté ou connaissant le contexte culturel, disposant du temps et des capacités émotionnelles nécessaires
- ◯ Soutien humain direct et en temps réel, par exemple une ligne d'assistance disponible sur plusieurs fuseaux horaires
- ◯ Outils algorithmiques dédiés pour faire respecter des normes locales spécifiques, y compris dans des langues minoritaires
- ◯ Hygiène numérique structurée, comme des périodes calmes ou nocturnes limitant, par exemple, l'envoi de messages à un par minute

SIGNAUX D'ALERTE

- ◯ Application incohérente des règles, notamment la censure d'idées diverses et non nuisibles
- ◯ Absence persistante de réponse aux préjudices, ou tolérance de ceux-ci au nom de la « liberté d'expression »
- ◯ Obligation de divulguer son identité réelle
- ◯ Sécurité « performative », par exemple la simple reproduction automatique de règles communautaires
- ◯ Infrastructure technique non protégée, exposée aux infiltrations ou aux fuites de données
- ◯ Dépendance aux algorithmes ou à des conditions d'utilisation génériques pour gérer les conflits interpersonnels

EDGELANDS

edgelands.institute