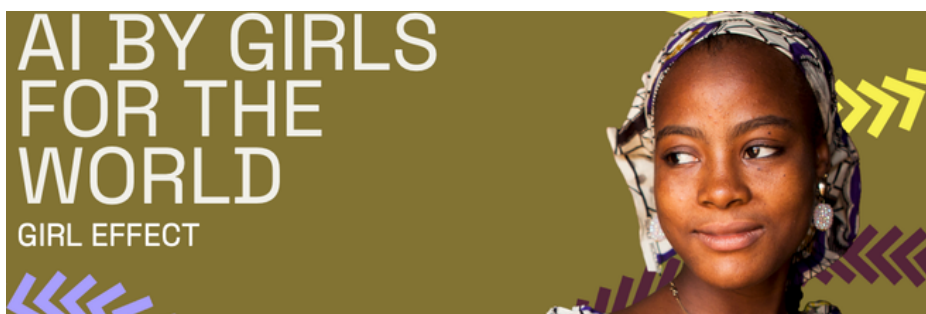




May 2026

Concevoir des chatbots d'IA éthiques — ce que cela exige en pratique

Séance 5 : Concevoir des chatbots d'IA éthiques – ce que cela exige en pratique



Notre cinquième séance de la série d'apprentissage Équité et IA portait sur ce qu'il faut réellement mettre en place pour concevoir des chatbots d'IA éthiques dans des contextes réels. Nous avons accueilli Karina Rios Michel, directrice de la création et de la technologie chez Girl Effect, une organisation mondiale à but non lucratif qui conçoit des médias et des outils numériques – y compris des chatbots d'IA – pour soutenir les adolescentes en Afrique et en Asie du Sud, rejoignant des millions d'utilisatrices. Leurs plateformes traitent maintenant des dizaines de milliers de messages par jour, ce qui représente un volume important d'interactions réelles et, par conséquent, une riche source d'apprentissage sur ce qui fonctionne, ce qui ne fonctionne pas et ce qu'il faut pour bâtir des systèmes d'IA réellement sûrs et utiles en pratique. Karina dirige le travail de Girl Effect à l'intersection de la technologie, de la créativité et de l'impact social, en mettant fortement l'accent sur la création d'outils ancrés dans les réalités des communautés desservies.

La séance était très pratique, avec de nombreuses questions réfléchies et beaucoup d'échanges tout au long de la discussion. Ce qui est ressorti le plus clairement, c'est le degré d'attention, de temps et d'intentionnalité qui sous-tend ce travail. Ce n'est pas un travail rapide ni facile. Il reflète un investissement à long terme dans la conception collaborative, l'apprentissage continu et l'itération. C'est cette profondeur d'effort qui a permis d'atteindre une véritable échelle et un impact réel, et qui a rendu les apprentissages partagés lors de cette séance si précieux pour d'autres organisations qui réfléchissent à la façon d'utiliser l'IA et les chatbots de manière significative et responsable.

Pourquoi cette séance est importante

Très peu d'organisations investissent le temps, le financement et les efforts à long terme nécessaires pour réellement concevoir et tester des outils d'IA dans des contextes réels – surtout de manière sûre, éthique et ancrée dans les réalités communautaires. Ce qui a rendu cette séance particulièrement précieuse, c'est que Girl Effect a précisément effectué ce travail et a choisi de partager ouvertement son parcours. Tout au long de la discussion, il était clair que ce travail a nécessité des années d'investissement : bâtir des systèmes, tester des modèles, tirer des leçons des échecs et s'adapter en continu. La séance a aussi mis en lumière un défi plus large dans le secteur : bien que de nombreuses organisations à but non lucratif explorent l'IA, beaucoup moins investissent dans la recherche, la documentation et le partage d'apprentissages de manière significative.

Karina a abordé directement cette tension. L'intérêt pour l'IA augmente, mais les apprentissages partagés demeurent limités, la recherche est fragmentée, et les pressions liées à la concurrence et à la protection de la propriété intellectuelle s'intensifient. Dans ce contexte, Girl Effect adopte une approche différente en priorisant l'ouverture, la collaboration et la contribution à une base de connaissances plus large sur ce qui fonctionne et ce qui ne fonctionne pas. Ce changement est important. Il crée un espace pour une utilisation plus ancrée et responsable de l'IA dans l'ensemble du secteur, en particulier pour les organisations qui travaillent dans des contextes complexes, aux ressources limitées et axés sur l'équité.

Ce que nous avons exploré ensemble : ce qu'il faut réellement pour concevoir un chatbot éthique

Karina a présenté l'ensemble du parcours de conception des chatbots de Girl Effect, depuis les premières expérimentations jusqu'au déploiement à grande échelle, ainsi que ce qu'il faut concrètement pour rendre ces outils utiles, sûrs et pertinents en pratique.

Quelques éléments sont ressortis clairement :

Ce travail commence bien avant l'IA

Le chatbot n'a pas commencé comme un chatbot. Il a évolué sur plus d'une décennie – à partir des plateformes de conseils de type magazine de Girl Effect (comme la rubrique « Baza Shangazi » (« Demande à tante ») du magazine Ni Nyampinga au Rwanda), puis vers les SMS et la radio, les plateformes numériques et, enfin, les chatbots propulsés par l'IA. Cet aspect est important parce que le fondement n'était pas la technologie : c'était l'écoute. Des années de questions posées par des filles et des jeunes. Une écoute portant sur les relations, la santé, l'identité et les défis du quotidien. Tout cela a façonné ce que le chatbot est finalement devenu. ***Le point essentiel ici est que, si l'on ne comprend pas en profondeur les questions que les personnes se posent réellement, l'IA ne comblera pas cette lacune : elle l'amplifiera.***

L'IA est une couche au sein d'un écosystème plus large

Les chatbots de Girl Effect, comme Big Sis et Bol Behen, ne sont pas des outils autonomes. Ils font partie d'un écosystème plus large qui comprend des marques médiatiques, du contenu social, la radio et l'engagement communautaire. Les jeunes découvrent généralement le chatbot par l'intermédiaire de ces canaux existants, dans des contenus auxquels ils et elles font déjà confiance et suivent, et avec lesquels ils et elles interagissent déjà. Le chatbot devient alors une continuité de cette relation, et non un point d'entrée nouveau ou inconnu. C'est important, car la confiance ne se construit pas uniquement par le chatbot. Elle se bâtit au fil du temps grâce à un contenu cohérent, pertinent et culturellement ancré. Le chatbot fonctionne parce qu'il s'inscrit dans cette confiance, et non à l'extérieur de celle-ci. Karina a souligné que, sans cet écosystème, les personnes utilisatrices pourraient ne pas interagir du tout, les questions sensibles pourraient ne pas être posées ou les réponses pourraient ne pas être jugées crédibles. Cela signifie que lancer un chatbot sans canal d'engagement existant – un auditoire, une plateforme ou une marque de confiance – pourrait limiter considérablement son utilité, en particulier auprès des jeunes. La question n'est donc pas seulement « comment concevoir un chatbot? », mais aussi « où cet outil s'inscrit-il dans la manière dont les personnes accèdent déjà à l'information et au soutien? » Cela modifie aussi notre façon de penser l'impact : la valeur ne réside pas seulement dans l'interaction avec le chatbot, mais dans ce qui se produit grâce à cette interaction – par exemple, si une personne apprend quelque chose de nouveau, prend une décision ou parvient à accéder à un soutien ou à des services auxquels elle n'aurait pas eu accès autrement. Dans le cas de Girl Effect, le chatbot représente souvent une étape dans un parcours plus long. Une conversation peut commencer par une question, mais elle peut mener à une compréhension plus approfondie, à un changement de comportement ou à une orientation vers du soutien concret. L'impact n'est pas l'interaction en soi : c'est ce que l'interaction rend possible.

Le soutien humain est intégré dès le départ

Une décision de conception centrale a été de garder des êtres humains dans la boucle dès le début. Karina a établi une distinction claire entre ce que l'IA peut bien faire et les limites qu'elle ne peut pas dépasser. L'IA est efficace pour répondre à des questions courantes, fournir de l'information à grande échelle et aider les personnes utilisatrices à comprendre leurs options. Mais lorsque les situations deviennent plus personnelles, émotionnelles ou à risque élevé, l'information seule ne suffit pas.



C'est là que le soutien humain devient essentiel. Girl Effect a intentionnellement conçu son système de manière à ce que les personnes utilisatrices puissent être dirigées vers des agentes et agents formés, capables d'offrir du réconfort, du contexte et de l'orientation d'une manière que l'IA ne peut pas offrir. Environ 30 % des interactions sont transférées à des modératrices et modérateurs ou à des conseillères et conseillers formés, en particulier dans tous les cas impliquant des abus, de la violence ou des divulgations sensibles. Une partie importante du travail a consisté à définir où l'IA est appropriée et où elle ne l'est pas. Cela comprend : ne pas générer certains types de conseils, ne pas répondre à des situations impliquant un risque ou un préjudice, et ne pas parler au nom d'une expérience vécue.

La co-conception est continue et profondément contextuelle

Ces chatbots sont co-conçus avec des jeunes et adaptés à chaque contexte, y compris à la langue, au ton, aux normes culturelles et même à l'humour. Karina a donné des exemples de modèles entraînés à comprendre des pratiques langagières locales, comme le sheng au Kenya, ou des messages qui combinent plusieurs langues, afin de veiller à ce que les réponses semblent naturelles plutôt que scénarisées ou extérieures au contexte. Sans ce niveau d'adaptation, les outils d'IA peuvent fonctionner sur le plan technique, mais ils ne parviendront pas à trouver écho, à bâtir la confiance ni à être utilisés.

Le passage à l'échelle repose sur des systèmes, et pas seulement sur la technologie

La capacité de Girl Effect à répondre à des dizaines de milliers de questions chaque jour ne tient pas seulement à l'utilisation de l'IA; elle reflète des années de travail pour bâtir des systèmes autour de cette technologie. Cela comprend des processus de modération et de protection, des voies de transfert vers du soutien humain, des partenariats avec des prestataires de services, ainsi que des essais et des ajustements continus du contenu et des réponses. Ce sont ces systèmes qui rendent possible l'utilisation sécuritaire de l'IA à grande échelle, en particulier dans des contextes qui touchent à des sujets sensibles comme la santé, les relations et la violence. Karina a souligné que cette infrastructure exige beaucoup de temps, d'investissement et de coordination. Ce n'est pas quelque chose que l'on peut reproduire rapidement en adoptant simplement un outil ou un modèle. Concevoir la technologie n'est qu'une partie du travail. Le travail le plus important, et la véritable responsabilité, résident dans la conception des systèmes, des partenariats et des mesures de protection qui rendent l'IA sûre et pertinente en pratique.

Principaux apprentissages : ce qui fonctionne en pratique

Plusieurs apprentissages clairs et pratiques se sont dégagés de l'expérience de Karina, particulièrement utiles pour les organisations qui réfléchissent à la pertinence de concevoir des outils d'IA dans leurs propres contextes, et à la manière de le faire.

- 1. Commencer par un besoin réel et à fort volume, et non par la technologie :** Girl Effect n'a pas commencé par « concevons un chatbot ». L'organisation recevait déjà un grand nombre de questions de la part de jeunes, dont plusieurs étaient profondément personnelles et difficiles à traiter à grande échelle. Le chatbot est né en réponse à cette pression : comment répondre à des milliers de vraies questions de manière rapide, cohérente et sûre? À retenir : l'IA est plus utile lorsqu'elle répond à une demande claire et existante, et non lorsqu'elle est introduite comme une nouvelle solution à la recherche d'un problème.
- 2. L'itération est constante et nécessaire :** le chatbot n'est pas un produit figé. Il est continuellement mis à l'essai, examiné et mis à jour à partir des interactions avec les personnes utilisatrices, des réponses signalées et de l'évolution des contextes et des besoins. Karina a expliqué que cela exige un investissement continu : des équipes doivent examiner les résultats, affiner les réponses et améliorer le système au fil du temps. À retenir : il ne s'agit pas d'une conception ponctuelle. C'est un processus continu qui exige des ressources importantes et constantes.

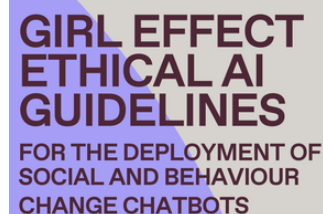
- 3. La confiance se construit par la cohérence, et pas seulement par l'exactitude :** l'exactitude est importante, mais il est tout aussi important que les réponses soient pertinentes, respectueuses et alignées avec la réalité de la personne utilisatrice. Cela passe par le ton et la langue, la pertinence culturelle et la cohérence entre les interactions. À retenir : *la confiance ne repose pas seulement sur la justesse de la réponse, mais aussi sur la manière dont cette réponse est vécue.*
- 4. Définir le succès en fonction des résultats, et non de l'utilisation du chatbot :** l'un des constats partagés portait sur la distinction entre l'engagement et l'impact. Un chatbot peut générer un grand volume d'interactions – des milliers de messages, des utilisateurs.rices récurrents.es, de longues conversations – sans pour autant atteindre des résultats significatifs. Girl Effect est très intentionnelle à ce sujet. L'objectif n'est pas de maximiser les conversations, mais de soutenir des progrès. Cela peut vouloir dire aider une personne à mieux comprendre une situation, à se sentir plus confiante dans une décision ou à faire un pas vers l'accès à du soutien. Dans bien des cas, le véritable impact se produit après l'interaction. À retenir : *le succès ne se mesure pas à l'intensité d'utilisation de l'outil, mais à ce qui change grâce à lui. Cela a des implications importantes pour les indicateurs d'utilisation et la façon d'évaluer l'impact.*
- 5. Concevoir en fonction de la manière dont les personnes veulent réellement interagir :** Karina a aussi expliqué que l'engagement ne découle pas simplement du fait d'offrir un chatbot; il vient d'une expérience conçue en fonction de la façon dont les personnes interagissent réellement. Plutôt que de miser uniquement sur un modèle conversationnel entièrement ouvert, Girl Effect a conçu une expérience hybride. Les personnes utilisatrices peuvent poser des questions directement, mais elles peuvent aussi naviguer dans du contenu structuré à partir de menus. Cela leur donne de la flexibilité : certaines veulent explorer, tandis que d'autres veulent poser une question précise. L'équipe a aussi accordé beaucoup d'attention au ton et à la facilité d'utilisation. Le chatbot est conçu pour être accessible et proche des pairs, et non clinique ou trop technique. Cela comprend l'utilisation d'un langage familier, de références culturellement pertinentes et d'une interface simple et intuitive. À retenir : *l'engagement se conçoit. Il découle de la compréhension des préférences des personnes utilisatrices et de la création d'une expérience qui semble naturelle, flexible et facile à utiliser.*
- 6. Être explicite sur ce que votre IA peut et ne peut pas faire :** un autre apprentissage important concerne l'ampleur des efforts déployés pour établir les attentes avec les personnes utilisatrices dès le départ. Girl Effect indique très clairement que la personne interagit avec un chatbot, ce que celui-ci peut et ne peut pas faire, ainsi que ses limites. Ce n'est pas seulement un détail technique : c'est un élément central de la conception éthique. La transparence aide à prévenir la dépendance excessive, réduit les risques de malentendu et permet d'éviter que les personnes utilisatrices accordent une confiance inappropriée au système. À retenir : *la clarté et la transparence sont essentielles à une utilisation sûre et responsable.*
- 7. Concevoir intentionnellement pour éviter la dépendance des personnes utilisatrices :** l'un des aspects les plus distinctifs de l'approche de Girl Effect est que le chatbot n'est pas conçu pour devenir un système de soutien émotionnel à long terme. Il est plutôt présenté comme un pont – un outil qui aide les personnes utilisatrices à comprendre leur situation et à se diriger vers du soutien concret. Cela se reflète dans la manière dont le chatbot communique, dans la structure des conversations et dans la façon dont les personnes sont encouragées à chercher de l'aide supplémentaire au besoin. En ce sens, Girl Effect n'optimise pas l'engagement de la même façon que plusieurs plateformes numériques d'IA. Le chatbot n'est pas conçu pour garder les personnes en conversation, mais pour les aider à avancer. Une interaction réussie peut signifier qu'une personne quitte le chatbot parce qu'elle se sent suffisamment informée pour agir, parler à quelqu'un ou accéder à un service. Cela reflète une conception très différente du succès : non pas la rétention, mais la progression. À retenir : *concevez votre outil pour soutenir l'autonomie, et non la dépendance.*

8. **Bâtir la sécurité au moyen de protections superposées, et non d'une seule fonctionnalité :** Karina a souligné que l'IA sécuritaire ne découle pas d'une seule fonctionnalité ou mesure de protection, mais de plusieurs couches qui fonctionnent ensemble. Cela comprend l'utilisation de contenus validés et approuvés par des spécialistes, l'établissement de limites quant à ce que l'IA peut générer, la surveillance des interactions afin de repérer les risques et la mise en place de voies claires de transfert vers du soutien humain. Ces couches sont continuellement mises à l'essai et ajustées en fonction des interactions réelles et des risques émergents. À retenir : *la sécurité doit être conçue dans l'ensemble du système, et non ajoutée à la fin.*
9. **Concevoir votre chatbot pour qu'il modèle un comportement respectueux, et pas seulement pour qu'il y réponde :** l'un des apprentissages plus inattendus partagés par Karina concernait le fait que certaines personnes interagissent parfois avec le chatbot en utilisant un langage abusif ou inapproprié. Plutôt que de simplement filtrer ou ignorer ces propos, Girl Effect a fait un choix de conception délibéré : le chatbot répond de manière à modéliser un comportement respectueux et à renforcer les limites. Au lieu d'obéir ou de poursuivre la conversation comme si de rien n'était, le chatbot peut recadrer la situation, indiquer clairement lorsqu'un langage n'est pas approprié et renforcer les normes liées au respect, au consentement et à la communication. Ce faisant, le chatbot devient plus qu'une simple source d'information : il fait partie de l'expérience d'apprentissage. Cela reflète un changement plus large dans la conception : il ne s'agit pas seulement d'optimiser la réactivité, mais aussi de façonner intentionnellement la manière dont les interactions se déroulent. Le chatbot est conçu pour refléter les valeurs que Girl Effect cherche à promouvoir : la confiance, le respect de soi et les limites saines. À retenir : *les outils d'IA ne sont pas neutres dans leur manière de répondre. Ils peuvent renforcer des dynamiques nuisibles ou, au contraire, modéliser activement les comportements et les valeurs que l'on souhaite encourager.*
10. **Planifier un investissement à long terme, et non des gains rapides :** l'une des réalités les plus importantes partagées pendant la séance est le niveau d'investissement derrière ce travail. Ce qui ressemble à un chatbot est en fait le résultat de plus d'une décennie de développement de programmes et de plusieurs années de travail consacré à l'IA. Cela comprend la création d'ensembles de données ancrés dans de vraies questions d'utilisatrices et utilisateurs, le développement et la mise à l'essai de contenu, l'établissement de partenariats avec des prestataires de services, des gouvernements et des communautés, ainsi que l'amélioration continue du système. À retenir : *des systèmes d'IA efficaces et éthiques exigent des efforts soutenus, de l'itération et un apprentissage au fil du temps.*
11. **Investir dans des données réelles et locales. Ne pas s'appuyer sur des modèles de langage d'IA génériques :** un élément essentiel de l'approche de Girl Effect est l'ampleur de son investissement dans la langue, la culture et le contexte. Plutôt que de s'appuyer sur des systèmes d'IA génériques et de traduire les résultats, l'organisation a développé sa propre couche d'IA entraînée sur des millions de messages anonymisés provenant de filles dans les pays où elle travaille. Cela permet au chatbot de refléter la manière dont les jeunes s'expriment réellement, y compris l'argot, les mélanges de langues, le ton et les références culturelles. En pratique, cela signifie que le chatbot ne se contente pas de traduire la langue; il comprend le contexte, les nuances et le sens d'une manière qui semble pertinente et digne de confiance. Ce niveau de pertinence ne vient pas de la technologie seule. Il découle d'années de :
- recueillir de vraies questions auprès des personnes utilisatrices;
 - bâtir des relations avec les communautés;
 - travailler avec des conseiller-e-s et spécialistes locaux;
 - adapter continuellement le contenu afin de refléter les réalités vécues.

Karina a souligné que, sans cet investissement, les outils d'IA risquent d'être fonctionnels sur le plan technique, mais culturellement non pertinents, voire trompeurs. À retenir : si votre IA n'est pas ancrée dans des données réelles et locales ainsi que dans l'expérience vécue, elle ne trouvera pas écho et pourrait causer plus de tort que de bien.

Ressources clés

Voici quelques ressources clés pour les personnes qui souhaitent explorer ce travail plus en profondeur :



- Girl Effect : aperçu du travail de Girl Effect dans la conception de médias et d'outils numériques – y compris des chatbots d'IA – visant à soutenir les adolescentes en Afrique et en Asie du Sud.
<https://www.girleffect.org/>
- Démo du chatbot d'IA (bac à sable multilingue) : un environnement de démonstration en direct partagé pendant la séance, qui permet d'explorer la performance de différents modèles et messages-guides dans diverses langues et différents contextes. Cela reflète le type d'environnement de test utilisé par les équipes de Girl Effect pour itérer et améliorer la conception des chatbots.
<https://genai.girleffect.org/chatbot/>
- Lignes directrices de Girl Effect sur l'IA éthique : un ensemble pratique de lignes directrices présentant l'approche de Girl Effect en matière de conception responsable de l'IA, y compris des principes liés à la sécurité, aux biais, à la protection et au travail auprès de populations vulnérables.
<https://merltech.org/wp-content/uploads/2025/04/Girl-Effect-Ethical-AI-Guidelines.pdf>
- Livre blanc de Girl Effect sur l'IA et l'apprentissage automatique : une analyse plus approfondie de son approche, y compris la conception du système, les décisions de mise en œuvre et les apprentissages tirés du déploiement de chatbots d'IA à grande échelle dans des contextes réels.
 - [Building with GenAI : chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://ge-prd-eu-west-1-webapp-assets-01.s3.eu-west-1.amazonaws.com/girleffect/media/documents/Girl-Effect_AI_ML-Whitepaper.pdf](https://efaidnbmnnnibpcajpcglclefindmkaj/https://ge-prd-eu-west-1-webapp-assets-01.s3.eu-west-1.amazonaws.com/girleffect/media/documents/Girl-Effect_AI_ML-Whitepaper.pdf)
 - [AI and ML Vision : chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://girleffect.org/static/theme/Girl_Effect_AI_MLvision_2024v2.pdf](https://efaidnbmnnnibpcajpcglclefindmkaj/https://girleffect.org/static/theme/Girl_Effect_AI_MLvision_2024v2.pdf)

Réflexion de clôture

Cette séance a été réellement inspirante, tant par la profondeur de l'expérience partagée que par l'ouverture dont les intervenantes ont fait preuve en expliquant ce qu'il faut concrètement pour concevoir une IA éthique en pratique. La participation a été forte tout au long de la séance, avec des questions et des réflexions pertinentes qui ont montré à quel point ce travail est à la fois important et complexe dans différents contextes. Ce qui ressort le plus, c'est que ce travail est possible. Mais il exige de l'intention, des investissements et une volonté d'apprendre et de s'adapter au fil du temps. Karina et l'équipe de Girl Effect ont offert un aperçu rare et précieux de ce à quoi cela ressemble réellement en pratique.

Si vous avez des idées de sujets pour de futures séances, des suggestions de conférencières ou conférenciers, des exemples de cas pertinents ou des questions que vous aimeriez que nous explorions, je serais ravie de vous lire. N'hésitez pas à communiquer avec moi, Paula Richardson, par courriel à richardson@salanga.org ou à me contacter sur LinkedIn : <https://www.linkedin.com/in/paula-nicole-richardson/>.