Global Coalition for Tech Justice

Proposition auprès du

Sommet pour l'action sur l'Intelligence Artificielle

10 et 11 février 2025, France

Groupe de travail sur

l'Intelligence Artificielle et l'intégrité de l'information

Global Coalition for Tech Justice

Novembre 2024

Thématiques :

IA de confiance

et

IA au service de l'intérêt public

Table des matières

Ces propositions ont été préparées par le Groupe de Travail sur l'IA du Global Coalition for Tech Justice.

Global Coalition for Tech Justice compte 250 membres, entre organisations et experts, provenant de 55 différents pays. Elle vise à garantir que les grandes entreprises technologiques jouent leur rôle dans la protection de la démocratie et des droits de l'homme dans le monde, et en particulier dans les pays de la Majorité Mondiale, là où les compagnies se sont souvent avérées négligentes dans leur gestion de l'impact des technologies ainsi que des plateformes déployées.

Les cosignataires individuels sont listés, toutefois, la proposition restera ouverte pour de nouvelles signatures jusqu'au mois de février 2025.

Digital Action est chargé de la convocation et de l'organisation du Global Coalition for Tech Justice. Depuis 2019, Digital Action mobilise un réseau mondial de partenaires dans le but d'exiger de meilleurs critères et normes de la part des gouvernements et des compagnies responsables pour nos environnements digitaux.

I. Introduction

De la conception au déploiement, de la formulation de politiques publiques à la reddition de comptes, il existe une profonde iniquité au sein du domaine de l'Intelligence Artificielle (IA), laquelle de plus en plus façonne l'avenir de l'intégrité de l'information, la démocratie et les droits de l'homme[1]. Cette iniquité représente un enjeu fondamental pour la justice à l'échelle mondiale et la gouvernance démocratique dans l'ère numérique. Au fur et à mesure que les systèmes d'IA s'enracinent au sein des institutions publiques clés et de l'infrastructure[2] - que ce soit dans les services de santé et d'éducation aux systèmes judiciaires et aux services publics - ce déséquilibre de pouvoir menace d'exacerber les disparités existantes à l'échelle mondiale et de créer de nouvelles formes de dépendance technologique qui compromettent l'autonomie nationale et l'autodétermination. L'accélération rapide du développement de l'IA, laquelle se trouve concentrée dans quelques centres de pouvoir[3] mondiaux, risque de rigidifier ces iniquités au sein de notre futur numérique partagé.

L'IA est principalement conçue dans les pays du Nord global ou en Chine, et déployée parmi toutes les régions, en prenant à peine en considération les contextes locaux aussi bien que les conséquences de ce déploiement. Les préjudices subis par la Majorité mondiale sont systématiquement ignorés, tandis que les capacités et l'expertise en matière d'élaboration de politiques publiques visant à encadrer l'IA pour garantir le respect des droits restent faibles dans ces régions. Ces préjudices vont du biais algorithmique[4] jusqu'au remplacement des systèmes locaux de prise de décision. Les impacts se manifestent de multiples façons : par exemple, par des systèmes d'IA qui ne reconnaissent pas les langues locales ni les nuances culturelles, par des systèmes de prise de décision automatisés entrainés avec des bases de données occidentales qui prennent des décisions inappropriées aux contextes du Sud mondial, ou encore par des systèmes de modération de contenu pilotés par l'IA qui suppriment par inadvertance des discours politiques légitimes[5]. Le manque d'expertise et de ressources

---

[1] United Nations, 'Il est urgent d'agir sur les risques que l'Intelligence Artificielle pose aux droits de l'homme' ('*Urgent Action Needed over Artificial Intelligence Risks to Human* Rights') (UN News, 17 Septembre 2021) https://news.un.org/en/story/2021/09/1099972, consulté le 5 Mai 2024.

[2] Les technologies d'usage général (TUG) 'sont des Technologies qui, au cours de l'histoire, ont changé l'ensemble de l'économie et, par conséquent, ont le potentiel de mettre en œuvre des changements drastiques dans la société avec un impact sur les structures économiques et sociales préexistantes'. André Guidetti, 'L'intelligence artificielle en tant que technologie à usage général : Une analyse empirique et appliquée de sa perception' ('*Artificial Intelligence as General Purpose Technology : An Empirical and Applied Analysis of its Perception'*) (Thèse de Master, Université de la Vallée d'Aoste, 2020), p.1 https://univda.unitesi.cineca.it/bitstream/20.500.14084/428/1/ETI_104_Guidetti_André.pdf consulté le 7 octobre 2024.

[3] Anu Bradford, 'La course à la régulation de l'intelligence artificielle' ('*The Race to Regulate Artificial Intelligence'*) (Foreign Affairs, 27 juin 2023) https://www.foreignaffairs.com/united-states/race-regulate-artificial-intelligence-sam-altman-anu-bradford, consulté le 25 octobre 2024

[4] Nations unies, 'L'impact des nouvelles technologies sur la promotion et la protection des droits de l'homme dans le contexte des rassemblements, y compris les manifestations pacifiques' ('*Impact of New Technologies on the Promotion and Protection of Human Rights in the Context of Assemblies, Including Peaceful Protests'*) (2020).

[5] Frederik Zuiderveen Borgesius, 'Discrimination, intelligence artificielle et prise de décision algorithmique' ('*Discrimination, Artificial Intelligence, and Algorithmic Decision-Making*') (Conseil de l'Europe, 2018). https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73, consulté le 8 October 2024.

locales pour identifier et traiter ces problèmes, combinée à l'exclusion systématique provoquée par les entreprises, des connaissances et de l'expertise du Sud mondial en matière de développement et de déploiement de l'IA génère un cycle auto-entretenu de dépendance technologique et de marginalisation.

Il existe une profonde asymétrie dans les capacités effectives d'application de la loi et dans les compétences juridictionnelles, qui fait que la plupart des pays de la Majorité Mondiale ne puissent pas participer effectivement à la conception des systèmes d'IA déployés dans leurs espaces d'information, même dans les cas où ils les auraient mis en place de régulations. Ce déséquilibre des pouvoirs compromet la souveraineté nationale et la gouvernance démocratique dans le domaine du numérique. Même lorsque les pays élaborent des réglementations intégrales en matière d'IA, ils font face à des difficultés considérables par rapport à la garantie d'application de ces règles face aux puissantes entreprises technologiques multinationales. La nature transnationale des systèmes d'IA, combinée à la concentration de l'expertise technique et juridique dans les pays du Nord, crée une situation dans laquelle les pays de la Majorité Mondiale doivent souvent accepter les systèmes et les politiques d'IA qui leur sont imposés, sans tenir compte des lois ou des normes sociales locales.

Les citoyens et la société civile sont marginalisés ou exclus à certaines étapes du processus – que ce soit lors de la conception, le déploiement, la reddition de comptes et l'élaboration des politiques publiques. Ils requièrent d'un accès et d'un soutien durable afin de renforcer leurs capacités et leur expertise pour pouvoir bénéficier d'une inclusion qui soit significative. Cette exclusion perpétue un cycle de dépendance technologique et de déficit démocratique. Les organisations de la société civile, qui traditionnellement jouent un rôle crucial dans la protection de l'intérêt public ainsi que dans la promotion de la participation démocratique, manquent souvent de l'expertise technique et des ressources nécessaires pour s'engager efficacement dans la gouvernance de l'IA. La complexité des systèmes d'IA et l'opacité délibérée de leurs processus de développement et de déploiement créent des barrières considérables qui empêchent une participation publique significative. Cette exclusion systématique des voix des citoyens signifie que les systèmes d'IA sont développés sans l'apport crucial des communautés qu'ils affecteront le plus.

C'est précisément la raison pour laquelle la transparence et l'explicabilité doivent être des principes fondamentaux du développement éthique de l'IA. Les citoyens ont le droit fondamental de comprendre comment les systèmes d'IA affectent leur vie et de recevoir des explications claires sur la prise de décisions automatisée[6]. La transparence remplit deux fonctions essentielles : elle permet au public de comprendre le fonctionnement des systèmes d'IA et, plus important encore, elle fournit les bases nécessaires pour tenir responsables les développeurs et les plateformes des impacts provoqués par leurs technologies. Sans la transparence, la participation citoyenne significative et la supervision efficace restent impossibles, ce qui accentue d'autant plus le déséquilibre de pouvoir entre les développeurs d'IA et les communautés impactées par leurs systèmes.

[6] Gabriela Arriagada Bruneau 'Les biais de l'algorithme : L'importance de concevoir une intelligence artificielle éthique et inclusive', (*'Los sesgos del algoritmo : La importancia de diseñar una inteligencia artificial ética e inclusiva'*) (La Pollera, 2024) https://lapollera.cl/libros/sesgos-algoritmo-ia-etica/ consulté le 28 octobre 2024.

L'une des conséquences de l'iniquité mondiale est l'intégration transversale de l'iniquité raciale dans la conception[7], le déploiement[8], l'accès à la reddition de comptes et l'élaboration des politiques. Ce biais systémique[9] aggrave les injustices sociales existantes et menace d'ancrer davantage les modèles historiques de discrimination. Le manque de diversité[10] dans les équipes de développement de l'IA, dans les données d'entraînement et dans les processus de test mène à des systèmes qui non seulement ne contribuent pas à combattre les iniquités raciales existantes, mais qui en plus, les renforcent de façon active. Que ce soient des systèmes de reconnaissance faciale qui s'avèrent peu performants sur les peaux foncées[11] ou des modèles de langage qui perpétuent des stéréotypes préjudiciels, les implications raciales des pratiques actuelles de développement de l'IA sont profondes et vastes. L'absence de mécanismes effectifs de reddition de comptes et d'analyse de risques fait que que ces biais restent souvent indétectés et ne soient pas corrigés jusqu'à ce que qu'un dommage important se soit produit.

Le Comité consultatif de haut niveau sur l'IA a établi des principes et recommandations clés pour une gouvernance mondiale de l'IA, en insistant sur le fait que ces principes ne peuvent être aboutis véritablement sans un affrontement aux iniquités fondamentales, en particulier celles qui séparent les pays du Nord mondial de ceux du Sud mondial. S'il est vrai que la vision inclusive et basée sur les droits promue par le Comité consultatif présente une base fondamentale, la transformation de ces principes en réalités concrètes requiert de mécanismes spécifiques et actionnables dans le sens d'une redistribution de pouvoir au sein de l'écosystème de l'IA. Notre proposition relative à *l'Intelligence Artificielle en tant que bien commun* offre une voie pratique pour la mise en place de ces principes, de façon à assurer que les voix historiquement mises à l'écart puissent participer de façon significative dans le développement et le déploiement de l'IA. Cette approche répond de façon directe aux inquiétudes exprimées par le Conseil consultatif concernant les brèches de représentation et les défis de mise en œuvre d'initiatives actuelles de gouvernance de l'IA, et offre des solutions concrètes pour la consolidation d'une réelle équité mondiale au sein de la gouvernance de l'IA.

---

[7] 'Chaque ensemble de données utilisé pour former des systèmes d'apprentissage automatique, que ce soit dans le contexte de l'apprentissage automatique supervisé ou non supervisé, qu'il soit considéré comme techniquement biaisé ou non, contient une vision du monde'. Kate Crawford, 'L'Atlas de l'IA : Pouvoir, politique et coûts planétaires de l'intelligence artificielle' *('The Atlas of AI : Power, Politics, and the Planetary Costs of Artificial Intelligence'*) (Yale University Press 2021) 139

[8] Joy Buolamwini et Timnit Gebru, 'Nuances de genre : Disparités intersectorielles de précision dans la classification commerciale du genre dans les actes de Machine Learning Research' ; Conférence sur l'équité, la responsabilité et la transparence (*'Gender Shades : Intersectional Accuracy Disparities in Commercial Gender Classification in Proceedings of Machine Learning Research' ; Conference on Fairness, Accountability, and Transparency)* (2018) 81:1-15 http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf consulté le 21 octobre 2024.

[9] Cathy O'Neil, 'Armes de destruction *en math*: Comment les Big Data augmentent les inégalités et menacent la démocratie' ('*Weapons of Math Destruction : How Big Data Increases Inequality and Threatens Democracy'*) (Crown 2016) 10

[10] Forum sur l'information et la démocratie, 'L'IA comme un bien public : Assurer le contrôle démocratique de l'IA dans l'espace d'information' *('AI as a Public Good : Ensuring Democratic Control of AI in the Information Space')*.https://informationdemocracy.org/wp-content/uploads/2024/03/ID-AI-as-a-Public-Good-Feb-2024.pdf. Consulté le 10 October 2024 p.24

[11] Morgan Meaker, 'Cet étudiant s'attaque à un logiciel d'examen biaisé' ('*This Student Is Taking On « Biased » Exam Software'*). (WIRED, 5 avril 2023) https://www.wired.com/story/student-exam-software-bias-proctorio/, consulté le 24 octobre 2024.

Dans le monde numérique d'aujourd'hui, l'intégrité de l'information consiste fondamentalement à autonomiser les citoyens à assumer le contrôle sur leurs informations - ce à quoi ils accèdent, ce qu'ils consomment, comment elles sont présentées et comment ils les évaluent. Néanmoins, nous sommes confrontés à un défi majeur : une poignée de grandes entreprises technologiques détiennent actuellement un pouvoir sans précédent sur notre écosystème de l'information, et ce sont elles qui déterminent ce que des milliards de personnes consomment, et la façon comment elles le consomment. Cette concentration de pouvoir n'est pas seulement une problématique dans le domaine commercial - elle influence aussi directement le discours démocratique, et l'affaiblit souvent.

L'intégrité de l'information requiert trois éléments essentiels : la transparence dans la manière dont l'information est sélectionnée et distribuée, la reddition de comptes de ceux qui contrôlent ces systèmes et une riche pluralité de sources d'information fiables. Si les Principes mondiaux des Nations Unies pour l'intégrité de l'information[12] constituent un cadre de référence important -en mettant l'accent sur la confiance sociale, les incitatifs sains, l'autonomisation du public, l'indépendance des médias et la transparence de la recherche- leur perspective centrée sur le Nord mondial requiert d'un examen critique. Du point de vue du Sud mondial[13], nous devons faire face à des défis structurels plus profonds: la souveraineté numérique face à la résistance aux mécanismes de gouvernance locale dont les plateformes font preuve ; la nécessité de promouvoir le journalisme comme une pratique éthique plutôt que de promouvoir simplement des « médias indépendants » ; et la reconstruction des espaces sociaux communs plutôt que de simplement renforcer la résilience face aux menaces extérieures. Il nous faut aller au-delà d'un monde dans lequel les plateformes de la Big Tech et leurs algorithmes dominent la curation d'information, tout en reconnaissant qu'une autonomisation significative de la part du public requiert que l'on adresse les inégalités sociales, raciales et de genre. Les États démocratiques doivent jouer un rôle actif dans la prévention de la concentration du marché et garantir un accès équitable aux données et aux opportunités de recherche, en particulier pour les chercheurs du Sud qui sont actuellement confrontés à des obstacles systémiques. Cette transformation nécessite non seulement d'innovations techniques et d'une supervision démocratique, mais aussi d'un engagement à affronter les inégalités structurelles qui façonnent notre écosystème de l'information.

Notre proposition vise à faire face à l'iniquité mondiale et à ses multiples composantes, notamment l'iniquité géographique, raciale et sociale, par une intervention coordonnée et systématique à de multiples niveaux. Cette approche globale reconnaît que la lutte contre l'iniquité en matière d'IA nécessite d'actions simultanées dans les domaines technique, social et politique. Elle nécessite de la création de nouvelles institutions et de nouveaux cadres de référence capables de redistribuer efficacement le pouvoir dans l'écosystème de l'IA, à la fois qu'elle vise à renforcer les capacités locales pour assurer une participation significative dans la gouvernance de l'IA. En abordant la nature interconnectée de ces inégalités, nous visons à

[12] Nations Unies, 'Principes mondiaux pour l'intégrité de l'information : Recommandations pour une action multipartite' ('*Global Principles for Information Integrity: Recommendations for Multi-stakeholder Action*') (2023) https://www.un.org/sites/un2.un.org/files/un-global-principles-for-information-integrity-en.pdf, consulté le 23 octobre 2024.

[13] Nina Santos, 'Cinq principes brésiliens pour l'intégrité de l'écosystème de l'information' *('Five Brazilian Principles for the Integrity of the Information Ecosystem'*) (Tech Policy Press, 2 novembre 2023) https://www.techpolicy.press/five-brazilian-principles-for-the-integrity-of-the-information-ecosystem/, consulté le 11 novembre 2024.

créer un changement durable qui permette aux communautés d'avoir une incidence sur les systèmes d'IA qui affectent leurs vies.

Il ne s'agit pas seulement de créer de nouvelles réglementations - mais de redistribuer fondamentalement le pouvoir dans l'espace de l'information numérique et de veiller à ce que la technologie serve la démocratie et les droits de l'homme au lieu de les affaiblir. À cette fin, nous proposons pour le Sommet Mondial sur l'IA, qui aura lieu en février 2025, les suivants résultats actionnables, qui ont été conçus dans la logique de générer un impact immédiat tout en renforçant la capacité à long terme de gouvernance démocratique des systèmes d'IA.

II. L'IA en tant que bien commun : la démocratisation de l'Intelligence Artificielle au travers du pouvoir citoyen à échelle mondiale

Nous proposons le lancement d'une initiative « L'IA en tant que bien commun » qui repose sur quatre piliers de mise en œuvre. Imaginez un réseau mondial où les citoyens, en particulier ceux qui n'ont jamais été pris part des négociations, puissent contribuer à façonner comment l'IA est conçue et utilisée. Au travers de ces quatre programmes interconnectés, nous pourrions mettre le pouvoir de l'IA dans les mains de tous : des centres de formation dans les pays du Sud ; des Conseils citoyens qui examinent les processus de conception et d'élaboration de politiques de l'IA ; des laboratoires où les personnes puissent expérimenter de nouvelles politiques en matière d'IA ; et un système intégral de supervision qui garantisse la reddition de comptes -. Chaque pilier joue un rôle essentiel : Le Réseau des Labos pour une Intelligence Artificielle Équitable renforce les capacités ; Le Conseil Citoyen de Conception permet une incidence directe de la part de la communauté ; Le Labo pour l'Innovation en Politiques Publiques relatives à l'Intelligence Artificielle permet une expérimentation en toute sécurité ; et le Système multi-acteur de supervision garantit que l'ensemble du processus se maintienne responsable et transparent. Il ne s'agit pas seulement de rendre l'IA plus juste ; il s'agit de s'assurer qu'elle serve à chacun d'entre nous, et pas seulement à quelques privilégiés. Il ne s'agit pas d'une vision lointaine de l'avenir- il est possible dés à présent de donner aux citoyens des outils concrets pour assurer que l'IA contribue à nos sociétés démocratiques au lieu de les entraver.

A. Le Réseau des Laboratoires pour une IA Équitable

Imaginez le Réseau des Laboratoires pour une IA Équitable comme une école mondiale pour les futurs responsables de la politique en matière d'Intelligence Artificielle. Or, cette école a une particularité : elle est spécialement conçue pour autonomiser les communautés qui n'ont traditionnellement jamais eu leur mot à dire sur le développement des technologies et les encourager à s'exprimer. Grâce à des centres physiques répartis dans les pays de la Majorité Mondiale ainsi qu'à une solide plateforme en ligne, le réseau crée des espaces où les citoyens peuvent acquérir une expérience pratique des systèmes d'IA, tout en apprenant comment orienter leur développement dans la bonne voie.

Ce qui rend ce réseau spécial est son approche intégrale. Il ne s'agit pas seulement de formation technique : les participants participent à un programme de bourses d'une durée d'un

an au cours duquel il leur est offert des enseignements sur toutes les thématiques allant de l'audit de l'équité des systèmes d'IA à l'élaboration de politiques qui protègent les intérêts de leurs communautés. D'ici à 2026, le réseau vise à former 1 000 nouveaux référents en matière de politiques publiques liées à l'IA en Afrique, en Asie, au Moyen-Orient et en Amérique Latine, de façon à renforcer leur compréhension des aspects techniques et sociaux de l'IA, pour ainsi créer une puissante force de changement positif dans le paysage mondial de l'IA.

Modèle de financement et de partenariat

La viabilité financière du Réseau des Laboratoires pour une IA Équitable repose sur un modèle de financement multipartite soigneusement conçu. Plutôt que de dépendre d'une seule source de financement, nous avons conçu une approche équilibrée qui répartit les ressources et les responsabilités entre différents secteurs. La participation des gouvernements des pays hôtes apporte un soutien infrastructurel et une légitimité qui sont essentiels. Leur investissement témoigne d'un engagement à développer l'expertise locale en matière de politique d'IA et garantit que le programme s'aligne aux objectifs de développement nationaux.

Les partenariats avec les entreprises apportent plus qu'un simple soutien financier. Les grandes entreprises technologiques fournissent des ressources techniques essentielles, des possibilités de mentorat et des études de cas réelles. Néanmoins, nous avons structuré ces partenariats de manière à préserver l'indépendance du réseau et sa capacité à évaluer les technologies de l'IA et leur impact de manière critique.

Les institutions internationales sont vitales pour garantir la pertinence et la durabilité du programme à l'échelle mondiale. Leur participation permet de maintenir des normes élevées et facilite le partage des connaissances entre les régions. Le modèle de financement comprend des mécanismes de viabilité à long terme, notamment un fonds de dotation et des activités génératrices de revenus qui soutiennent la croissance du réseau tout en préservant sa mission centrale.

Parcours éducatif et structure de formation

Le parcours éducatif proposé par le Réseau des Laboratoires pour une IA Équitable représente une approche visant à combler l'écart entre l'expertise technique et la compréhension des politiques publiques en matière de gouvernance de l'IA. Fondamentalement, le programme reconnaît que pour que les référents en matière de politiques publiques liées à l'IA soient efficaces, ils ont besoin d'une compréhension globale qui couvre à la fois les dimensions techniques et sociales de celle-ci. Pour y parvenir, nous avons mis au point un système sophistiqué avec deux filières, qui s'adapte aux antécédents des participants tout en garantissant que chacun puisse acquérir un ensemble de compétences pluridisciplinaires.

La filière surnommée « Les systèmes d'IA et la conception de politiques publiques », destiné aux professionnels de la politique publique et du droit, commence par démystifier la technologie de l'IA. Les participants commencent par une expérience pratique des concepts de base de la programmation et de l'apprentissage automatique, passant de la compréhension théorique à l'application pratique. Par le biais de laboratoires interactifs et de projets concrets,

ils apprennent à évaluer les systèmes d'IA de manière critique, à comprendre leurs limites et à évaluer leur impact sur la société. À la fin du programme, ces participants peuvent communiquer efficacement avec des équipes techniques et prendre des décisions politiques éclairées sur la base d'une véritable compréhension technique.

D'autre part, les professionnels techniques qui intègrent le programme suivent la filière « Intégration des politiques, des droits de l'homme et de l'éthique », transformant leur expertise technique en connaissances pertinentes pour les politiques publiques. Ce parcours met l'accent sur le paysage réglementaire, les cadre de référence internationaux des droits numériques et les impacts plus nuancées de l'IA sur différentes communautés. Les participants apprennent à traduire leur expertise technique en recommandations pour les politiques publiques, tout en prenant en considération les divers contextes culturels et besoins sociétaux.

Les deux filières convergent vers un tronc commun qui développe des compétences cruciales en matière de *leadership* et de plaidoyer. Cette expérience partagée crée un puissant réseau de professionnels qui peuvent construire des ponts entre les domaines techniques et politiques, traditionnellement très écartés entre eux.

Sélection et répartition régionale

Le processus de sélection pour le Réseau des Laboratoires pour une IA Équitable est conçu pour construire une communauté diversifiée et influente de futurs référents en politiques publiques en matière d'IA. Conscients que les différentes régions font face à des défis spécifiques et ont accès à des opportunités différenciées en matière de développement de l'IA, nous avons mis en place un système de quotas équilibré qui garantit la représentativité parmi les régions de la Majorité Mondiale. Il ne s'agit pas seulement d'une question de chiffres, mais de la création d'un dialogue riche entre différentes perspectives et expériences.

Nos critères de sélection vont au-delà des métriques traditionnelles. Si bien que l'expérience professionnelle est importante, nous apprécions particulièrement les candidats qui font preuve d'une profonde compréhension de leur contexte régional et qui ont le potentiel de catalyser le changement au sein de leur communauté. Nous recherchons des personnes capables de faire le lien entre les développements mondiaux en matière d'intelligence artificielle et les besoins locaux, en tenant compte de facteurs tels que leur engagement dans des initiatives communautaires et leur capacité à gérer des relations complexes avec les différents acteurs.

La structure des pôles physiques est fondamentale pour mener à bout notre vision. Chaque pôle, qu'il soit situé à Nairobi, à Jakarta ou à São Paulo, sert de centre d'excellence pour sa région, adapté aux contextes locaux tout en remplissant des critères mondiaux. Ces pôles ne sont pas de simples centres de formation; ce sont des incubateurs pour l'innovation régionale en matière de politique publique d'IA, conçus pour favoriser la collaboration entre les participants et les parties prenantes locales.

Cadre de référence de mise en œuvre et gouvernance

La stratégie de mise en œuvre reflète notre engagement dans la construction d'une institution durable, capable de s'adapter et de se développer. Notre structure de gouvernance associe une supervision globale à une autonomie régionale, ce qui garantit que les programmes restent adaptés aux contextes locaux tout en respectant les normes internationales. Le Conseil Consultatif International joue un rôle crucial dans la direction stratégique, en apportant des perspectives diverses d'experts techniques, de spécialistes en politiques publiques et de référents de la société civile.

L'assurance de qualité est transversale à tous les aspects du programme. Des révisions régulières du parcours éducatif, des mécanismes de retour d'informations de la part des participants, ainsi que des audits externes garantissent que le réseau continue à répondre aux besoins évolutifs de l'élaboration des politiques publiques en matière IA. L'évaluation de l'impact va au-delà des mesures traditionnelles, en évaluant la manière comment les participants influencent la politique d'IA dans leur région et créent un changement positif dans leur communauté.

Le calendrier de mise en œuvre est ambitieux mais réaliste. Commencer par des régions pilotes nous permet d'affiner notre approche avant de monter en échelle. D'ici 2026, notre objectif de former 1 000 référents en matière de politiques publiques liées à l'IA n'est pas uniquement une question de chiffres - il s'agit de créer une masse critique d'expertise dans des régions qui ont toujours été sous-représentées dans les discussions sur la gouvernance mondiale de l'IA.

B. Conseil Citoyen de Conception (CCC)

Le Conseil Citoyen de Conception révolutionne la manière dont l'intelligence artificielle est conçue, en associant des citoyens non-spécialistes au processus de développement de celle-ci. Avec des pôles régionaux situés en Afrique, en Amérique latine, en Asie et au Moyen-Orient, il crée une structure où les communautés - en particulier celles qui sont souvent négligées dans le développement des technologies - ont réellement leur mot à dire sur la façon dont les systèmes d'IA sont construits et déployés dans leurs régions.

Il ne s'agit pas seulement d'un retour d'informations à posteriori : le CCC implique les communautés à chaque étape du développement de l'IA. Avant de construire un système, il évalue son impact culturel et les besoins de la communauté. Pendant le développement, il peut tester des prototypes et faire un suivi de leurs effets. Et après le déploiement, il vérifie en permanence l'impact de ces systèmes sur la vie des communautés. Il s'agit de s'assurer que l'IA fonctionne pour tous, et pas seulement pour quelques experts en technologie.

C. Le Laboratoire pour l'Innovation en Politiques Publiques relatives à l'Intelligence Artificielle (LIPPIA)

Imaginez un espace où une diversité de voix – allant de décideurs politiques aux activistes de la société civile en passant par les communautés affectées, les défenseurs des droits de l'homme, les universitaires et les entités privées - puissent collaborer pour voir et expérimenter comment les politiques d'IA affectent les droits de l'homme fondamentaux et la vie quotidienne des communautés avant leur mise en œuvre. C'est ce qu'offre Le Laboratoire pour l'Innovation en

Politiques Publiques relatives à l'Intelligence Artificielle. Ses outils de visualisation de pointe et ses espaces de simulation rassemblent des technologues, des experts en droits de l'homme, des référents communautaires, des représentants d'entreprises et des décideurs politiques afin de transformer des idées abstraites de politique publique en scénarios tangibles qui démontrent les impacts réels sur protection la vie privée, la liberté d'expression, la non-discrimination et d'autres droits de l'homme essentiels.

Le laboratoire associe des outils de haute technologie à une approximation de l'élaboration de politiques publiques centrée sur les droits de l'homme, grâce à une approche innovante et multi-acteur. Dans ses stations de travail collaboratives et ses salles de réalité virtuelle, des référents de groupes indigènes travaillent côte à côte de défenseurs des droits numériques ; des organisations de base communautaire établissent des partenariats avec des fonctionnaires des institutions gouvernementales ; et des chercheurs académiques s'associent à des représentants de la jeunesse pour tester la manière dont différentes politiques d'IA pourraient avoir un impact sur les communautés vulnérables et les libertés fondamentales. Grâce à des simulations immersives, ce groupe diversifié peut expérimenter directement comment les décisions peuvent avoir un impact sur l'accès à l'information, la protection de la vie privée ou l'amplification de la discrimination. Une attention particulière est accordée aux impacts intersectionnels, ce pourquoi ce sont les communautés concernées qui mènent la conversation sur comment les politiques publiques peuvent avoir un impact différentié sur les personnes en fonction de leur genre, leur appartenance ethnique, leur statut économique ou leur situation géographique. C'est comme si l'on disposait d'un terrain de jeu pour les politiques publiques, doté d'une conscience et d'une sagesse collective, dans lequel les idées peuvent être testées et affinées en toute sécurité grâce à des perspectives multiples afin de garantir qu'elles protègent et renforcent les droits de l'homme avant d'avoir une incidence sur des millions de vies. L'approche participative du Laboratoire garantit que les politiques ne soient pas seulement créées pour les communautés, mais avec les communautés, ce qui permet d'éviter des conséquences imprévues, susceptibles de compromettre la dignité humaine ou d'exacerber les inégalités existantes.

D. Le Système multi-acteur de supervision (SMAS)

Le paysage actuel de l'IA présente un paradoxe critique. Alors que des entreprises comme OpenAI, Meta et les géants régionaux de la technologie déploient rapidement des systèmes d'IA à l'échelle mondiale, les mécanismes de supervision restent fragmentés et souvent inefficaces. Nous en avons été témoins par des contrastes frappants : le contrôle strict de la Chine sur l'accès à ChatGPT, par rapport à l'approche largement autorégulatrice des États-Unis, ou par rapport encore à l'approche européenne d'une réglementation centrée sur les droits, par rapport aux cadres de référence émergents dans les régions en développement (limités par les défis d'équité décrits ci-dessus). Ces disparités mettent en évidence le besoin urgent d'un système de supervision équilibré et adaptable qui permette de faire le lien entre ces approches, tout en priorisant les droits de l'homme et l'intégrité de l'information.

La composition de ces organes est soigneusement équilibrée. Les experts techniques travaillent aux côtés des défenseurs des droits de l'homme ; les spécialistes en Droit collaborent avec les journalistes ; et les représentants de la société civile veillent à ce que les voix des communautés restent au cœur de toutes les décisions. Cette diversité n'est pas seulement une

question de représentation : il s'agit de réunir les compétences nécessaires pour comprendre à la fois les implications techniques des systèmes d'IA et leur impact réel sur les communautés.

Le SMAS reconnaît que la gouvernance de l'IA est confrontée à des défis différents selon les régions. En Ouganda, où le gouvernement est en train de réviser sa stratégie en matière d'IA, le SMAS pourrait fournir un cadre de référence pour une supervision efficace tout en soutenant l'innovation locale. Dans des régions comme l'Asie du Sud-Est, où la localisation des données et les intérêts des États jouent un rôle important, il pourrait offrir des mécanismes flexibles qui respectent la souveraineté, tout en garantissant la protection des droits de l'homme.

La capacité du SMAS de s'adapter tout en conservant ses principes fondamentaux le rend particulièrement puissant. Dans les régions où l'autorégulation domine, il fournit des mécanismes de supervision structurés. Toutefois, dans les régions où le contrôle de l'État est fort, il offre des canaux pour la participation de la communauté et la protection des droits. Cette capacité d'adaptation assure que le SMAS reste pertinent et efficace dans des environnements réglementaires différents entre eux.

Prenons un exemple concret : Lorsqu'une grande entreprise d'IA souhaite déployer un nouveau modèle linguistique en Afrique de l'Ouest, l'Organisme Régional de Supervision évalue non seulement les spécifications techniques, mais aussi les implications culturelles, les préoccupations en matière de confidentialité des données et les impacts potentiels sur les écosystèmes d'information locaux. Il ne s'agit pas d'un simple exercice de vérification, mais d'une évaluation complète qui peut conduire à des modifications significatives, voire à des restrictions de déploiement si nécessaire.

Le SMAS garantit que le développement et le déploiement de l'IA restent transparents et responsables vis-à-vis de toutes les communautés qu'ils affectent, et se focalisent en particulier sur la protection des droits de l'homme et l'intégrité de l'information. Elle crée un cadre de référence intégral à l’échelle mondiale dans laquelle les l'Organismes Régionaux de Supervision, qui sont composés de représentants locaux, aussi bien que de défenseurs des droits de l'homme, de journalistes, de vérificateurs de faits, d'organisations indépendantes de médias, de membres de la société civile et des communautés concernées, ont un réel pouvoir de faire un suivi, d'évaluer et d'influencer la manière par laquelle les systèmes d'IA influencent les droits de l'homme et l'intégrité de l'information dans leur région.

Le SMAS supervisera les systèmes d'IA, aussi bien ceux déjà déployés que ceux en étape de pré-déploiement, qui ont un impact significatif sur les flux d'information et les droits de l'homme dans la société, en mettant l'accent sur quatre domaines critiques : (1) les Grands modèles de langage et les systèmes d'IA générative qui influencent le discours public et la création d'informations, (2) les systèmes de recommandation et de modération de contenu qui façonnent la distribution et l'accès à l'information, (3) les systèmes de prise de décision automatisés qui affectent les droits fondamentaux et les services publics, et (4) les technologies de surveillance et de suivi alimentées par l'IA. Cette supervision comprend une évaluation complète de l'impact de ces systèmes sur l'intégrité de l'information, y compris leur rôle dans l'amplification ou l'atténuation de la désinformation, leurs effets sur le pluralisme des médias et leur influence sur les biais algorithmiques dans la distribution de l'information. En parallèle, elle conserve la rigueur en ce qui concerne la reddition de comptes en matière de droits de l'homme, par le biais d'évaluations d'impact obligatoires, de mécanismes de réclamation

contraignants et de programmes de suivi dirigés par les communautés. Le SMAS utilise une double approche: d'une part, une supervision proactive par le biais d'évaluations préalables au déploiement et de l'autre, un suivi continu par le biais d'évaluations postérieures au déploiement, en accordant une attention particulière aux systèmes déployés par les deux plus grandes entreprises technologiques ainsi que par les entités gouvernementales. Le processus de supervision est ancré dans des exigences de documentation transparentes, des audiences publiques régulières et des mécanismes d'application clairs, garantissant ainsi que le développement et le déploiement de l'IA restent redevables face aux communautés affectées, tout en protégeant l'intégrité de l'information et les droits de l'homme.

Le SMAS transforme les mécanismes de supervision grâce à cinq approches clés, alignées aux Principes mondiaux des Nations Unies pour l'intégrité de l'information[14] :

1. **La construction de relations de confiance** : Établir des mécanismes de vérification pour les contenus générés par l'IA et promouvoir la transparence dans les systèmes algorithmiques.
2. **La restructuration des incitations** : Faire passer les plateformes d'un système de mesure basé sur l'engagement à un système de mesure de la qualité de l'information.
3. **L'autonomisation du public :** Promouvoir la culture et l'alphabétisation numérique et créer des outils pour la supervision publique des systèmes d'IA
4. **La protection des médias** : Sauvegarder l'indépendance et la diversité journalistiques dans l'ère de l'IA
5. **L'accès à la recherche** : Veiller à ce que les chercheurs puissent étudier de manière significative l'impact de l'IA sur les écosystèmes d'information.

Grâce à son approche intégrée qui met « les droits d'abord », le SMAS veille à ce que les systèmes d'IA respectent les droits de l'homme et contribuent à un environnement d'information sain, fiable et divers. Le cadre de référence de reddition de comptes comprend des mécanismes d'application allant d'alertes publiques et d'amendes, jusqu'à des restrictions d'exploitation en cas de violations graves des normes relatives aux droits de l'homme ou des principes d'intégrité de l'information.

- Un nouveau modèle de supervision démocratique de l'IA

Le SMAS représente une approche de la supervision de l'IA distinctement positionnée entre les observatoires de la société civile et les organismes de réglementation officiels. Contrairement aux tribunaux, qui prennent des décisions contraignantes, ou aux régulateurs, qui appliquent des règles, le SMAS sert d'observatoire transparent, voué à mettre la lumière sur la façon comment les systèmes d'IA affectent notre écosystème d'information et les droits de l'homme. Son pouvoir ne réside pas dans l'application de la loi, mais dans sa capacité à

[14] Nations Unies, 'Principes mondiaux pour l'intégrité de l'information : Recommandations pour une action multi-acteur' (*'Global Principles for Information Integrity: Recommendations for Multi-stakeholder Action'*) (2023) https://www.un.org/sites/un2.un.org/files/un-global-principles-for-information-integrity-en.pdf, consulté le 23 October 2024.

rassembler des preuves, à mettre en évidence des motifs et à permettre un discours public éclairé sur l'impact sociétal de l'IA. Le rôle d'observatoire du SMAS s'aligne sur les modèles émergents de supervision non réglementaire de l'IA, similaires au mécanisme de *Certification volontaire pour l'IA d'intérêt public* qui s'est avéré efficace dans d'autres domaines. Comme le souligne la recherche du Forum sur l'Information et la Démocratie[15], de tels mécanismes peuvent contribuer à remédier aux asymétries d'information entre les fournisseurs d'IA et le public, tout en créant des incitations positives en faveur d'un développement responsable. Similaire aux organismes de certification performants qui conservent leur indépendance vis-à-vis de l'industrie et du gouvernement tout en favorisant la transparence et la reddition de comptes, le SMAS peut servir comme un intermédiaire de confiance, qui permette un engagement significatif des parties prenantes et une mise à l'examen de la part du public. Sa position en tant qu'observatoire indépendant lui permet de documenter et d'analyser les impacts sociétaux des systèmes d'IA, tout en évitant les risques de capture réglementaire ou d'influence politique qui peuvent affecter des organismes de supervision plus formels. Cette approche permet au SMAS de renforcer la confiance et la légitimité, au travers d'une évaluation rigoureuse et d'une documentation publique plutôt que par des pouvoirs d'application.

- La légitimité démocratique par la mise en place de structures et de processus

Au fond, la légitimité du SMAS découle de sa structure profondément démocratique. Le *leadership* est assuré à tour de rôle par des représentants de différentes régions et de différents groupes de parties prenantes, avec une stricte limitation des temps de mandats, qui empêchant toute perspective unique de dominer. Un processus de nomination transparent garantit une représentation diversifiée, tandis que des politiques claires en matière de conflits d'intérêts ainsi que la multiplicité de sources de financement protègent contre l'accaparement par des intérêts puissants. Des audits indépendants réguliers sur les propres opérations du SMAS garantissent qu'il pratique la transparence qu'il préconise dans les systèmes d'IA.

- Construire la fiabilité au travers de la rigueur méthodologique

Les évaluations du SMAS gagnent en crédibilité grâce à leur rigueur méthodologique plutôt qu'à leur autorité réglementaire. Ses cadres de référence d'évaluation sont issus de consultations publiques et font l'objet d'une évaluation par les pairs, ce qui permet de s'assurer qu'ils reflètent des perspectives diverses et les recherches actuelles. Avant de publier des conclusions, plusieurs équipes indépendantes doivent vérifier les résultats, et la documentation complète de leurs méthodes est rendue publique. Cette approche permet d'obtenir des informations fiables sur l'impact sociétal des systèmes d'IA, tout en dressant des limites claires entre l'observation et la prescription.

- Mesures de sauvegarde contre le détournement de mission et la concentration de pouvoir

[15] Forum sur l'information et la démocratie, 'Un mécanisme de certification volontaire pour l'IA d'intérêt public: étude de la conception et des spécifications d'institutions mondiales dignes de confiance pour gouverner l'IA' (*'A Voluntary Certification Mechanism for Public Interest AI: Exploring the Design and Spécifications of Trustworthy Global Institutions to Govern AI'*) (Document de recherche, septembre 2024).

Pour éviter que le SMAS ne devienne *de facto* un censeur ni un système judiciaire parallèle, sa charte interdit de façon explicite les pouvoirs de modération de contenu et limite son champ d'action aux problèmes systémiques plutôt qu'aux cas individuels. Des audits externes réguliers évaluent le respect de ces limites, tandis que des rapports de transparence obligatoires détaillent ses activités et ses processus de prise de décision. Un solide système de protection des lanceurs d'alerte encourage la reddition de comptes interne, garantissant que SMAS reste fidèle à sa mission.

- Une approche collaborative de la supervision

Plutôt que de travailler de manière isolée, le SMAS collabore activement avec les institutions existantes tout en respectant leurs rôles distincts. Il fournit des éléments probants et des perspectives aux tribunaux et aux régulateurs, sans pour autant tenter de reproduire leurs fonctions. Ses évaluations soutiennent l'élaboration de politiques publiques éclairées et le débat public, sans prescrire de solutions spécifiques. Cette approche collaborative renforce la supervision démocratique des systèmes d'IA tout en préservant la séparation des pouvoirs, essentielle à la gouvernance démocratique.

- Autonomiser la compréhension et la construction d’un discours public

L'impact ultime du SMAS découle de sa capacité à éclairer des questions complexes pour la compréhension du public. Grâce à des réunions publiques régulières, à des séances de retour d'informations de la part de la communauté et à des données ouvertes sur ses opérations, il aide les citoyens à mieux comprendre les questions relatives au rôle de l'IA dans la société et à s'engager dans la thématique. Plutôt que de prendre des décisions à la place du public, il autonomise ce dernier à participer de façon informée à des débats cruciaux sur la manière dont l'IA façonne notre environnement d'information et nos processus démocratiques.

Cette approche soigneusement équilibrée garantit que le SMAS enrichisse la supervision démocratique des systèmes d'IA sans tomber dans la censure ni empiéter sur le territoire de la Justice. Le fait de maintenir des limites claires tout en fournissant des clés de compréhension cruciales renforce les institutions démocratiques existantes au lieu de prendre leur place.

III. Remarques finales

L'initiative « L'Intelligence Artificielle en tant que bien commun » représente une vision pour la démocratisation de la gouvernance de l'IA au travers de quatre piliers interconnectés, qui abordent à la fois les besoins immédiats et les changements structurels à long terme. En conjuguant les efforts de renforcement des capacités au travers du Réseau des Labos pour une Intelligence Artificielle Équitable, avec les contributions de la communauté au travers du Conseil Citoyen de Conception, l'expérimentation en politiques publiques au sein du Laboratoire pour l'Innovation en Politiques Publiques relatives à l'Intelligence Artificielle et une supervision transparente au travers du Système multi-acteur de supervision, ces cadres de référence créent de multiples voies pour une participation publique significative, qui puisse avoir une réelle incidence sur le façonnement du développement de l'intelligence artificielle.

Il est important de remarquer que cette structure se centre sur des perspectives promues par les voix du Sud global. Elle aborde également les iniquités de pouvoir fondamentales au sein de l'écosystème d'acteurs actuel de l'IA, tout en promouvant la consolidation de l'expertise et de la capacité de prise de décisions à l'échelle locale. Cette démarche intégrale reconnaît que la gouvernance effective de l'IA requiert à la fois d'innovations technologiques et de transformations sociales – allant de l'affrontement des inégalités raciales et géographiques, à la garantie de la souveraineté digitale, et à la reconstruction d'espaces sociaux partagés. Le succès de cette initiative dépendra d'un engagement durable et soutenu envers une gouvernance inclusive, de la transparence dans les processus et d'une véritable redistribution de pouvoir, afin de garantir que l'IA puisse, non pas compromettre, sinon servir aux valeurs démocratiques et aux droits de l'homme.

Cette proposition est ouverte à la signature des organisations et des individus jusqu'au 10 février 2025. La liste de signataires sera publiée régulièrement.

Pour signer la proposition, vous êtes invités à remplir ce formulaire.