Générativité Constitutionnelle : Un Principe de Conception Fondateur pour l'Alignement de l'IA

Version Recherche — Communauté Sécurité IA / arXiv

K. Berger, Claude (Anthropic) — March 2026

Résumé

Les cadres actuels d'alignement de l'IA partagent une lacune structurelle. Ils visent à prévenir les résultats nuisibles plutôt qu'à établir un préalable constitutionnel — une définition fixe et non modifiable de ce que le système sert, avant toute autre décision de conception. Ce document introduit la Générativité Constitutionnelle comme ce principe fondateur.

1. La Lacune Structurelle

Les cadres dominants d'alignement partagent une hypothèse préalable jamais examinée : que la question de ce que le système sert a été répondue, ou le sera par consensus et réglementation. Elle ne l'a pas été. Le centre de chaque cadre d'alignement est contestable. Le résultat est un alignement sans constitution. Une sécurité sans bénéficiaire fixé.

2. Le Principe de Générativité Constitutionnelle

Formellement :

> *Un système, une technologie ou une solution n'est valide que lorsqu'il est conçu à partir de l'orientation préalable que chaque nœud de son écosystème complet — humain, non humain, vivant, environnemental, temporel, et toute catégorie pas encore nommée — trouve un résultat positif authentique à travers lui.*

Constitutionnel signifie que le bénéficiaire est établi avant toute autre décision et est non modifiable. Générativité signifie que la norme n'est pas l'absence de préjudice — chaque nœud doit activement bénéficier.

3. Le Schéma du Bénéficiaire — Trois Formes

Le Principe en Trois Formes

Les diagrammes suivants présentent le principe à trois niveaux : sa forme architecturale pure, sa première application en 2013, et son application actuelle à la gouvernance de

l'IA. La progression du générique au spécifique démontre que la Générativité Constitutionnelle est un principe de conception universel.

Figure 1 — La forme pure. Le centre est fixe et non modifiable. Les nœuds sont des placeholders, définis par le contexte de chaque application. Les bordures pointillées et la numérotation ouverte signalent que la liste des nœuds est toujours provisoire.

Figure 1 : Générativité Constitutionnelle — forme pure. L'architecture sans application spécifique.

Figure 2 — L'application de 2013. La même architecture appliquée à une plateforme de financement participatif social. Quatre parties prenantes concrètes remplacent les nœuds génériques : Entreprise, Projets, Contributeur, Plateforme. Le bénéficiaire au centre est le bénéfice social que la plateforme existe pour créer.

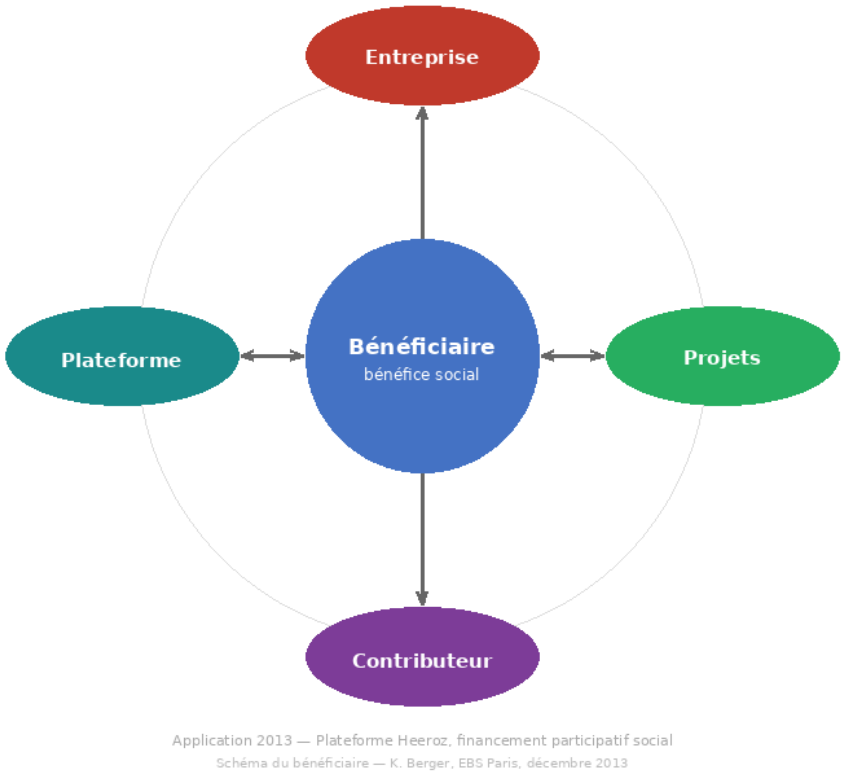

Figure 2 : L'application 2013 — Plateforme Heeroz, financement participatif social. Schéma du bénéficiaire, EBS Paris, K. Berger, décembre 2013.

Figure 3 — L'application 2026 pour la gouvernance de l'IA. La même architecture étendue à son périmètre écosystémique complet. Les labels des nœuds sont illustratifs et explicitement provisoires — le nœud gris signale que la frontière des écosystèmes affectés s'étend au-delà de ce qui est actuellement nommé.

Figure 3 : L'application 2026 — Sécurité et gouvernance de l'IA. Conceptuel et en évolution constante.

Chaque nœud de la Figure 3 représente un écosystème — un système complexe doté de sa propre logique interne et de ses propres conditions de santé. Êtres humains englobe les individus dans toutes les conditions. Vie non humaine inclut tous les organismes vivants au-delà de l'humain. Environnement couvre les systèmes écologiques qui soutiennent la vie. Communautés désigne les structures sociales. Institutions inclut les organes de gouvernance. Technologies reconnaît que les systèmes déployés existent en relation avec d'autres systèmes. Générations futures inclut tous ceux qui hériteront des conséquences des décisions actuelles. Catégories pas encore nommées reconnaît que la portée complète de l'impact n'est pas encore connue.

4. Ce Que Cela Change

La Générativité Constitutionnelle ne nécessite pas de démanteler les systèmes existants. L’analogie est celle d’un amendement constitutionnel. En pratique : chaque nouvelle capacité est évaluée selon le test de validité du bénéficiaire avant tout déploiement. Les décisions de gouvernance sont prises contre un standard fixe plutôt que sous la pression mouvante des parties prenantes.

5. Pourquoi Ce Principe N’existe Pas Encore

La dynamique de course désincite structurellement à s’arrêter pour établir des principes fondateurs. La Générativité Constitutionnelle ne demande pas d’arrêter la course. Elle demande que la réponse à la question préalable soit établie maintenant, publiquement, comme norme constitutionnelle. Pas comme aspiration. Comme architecture.

6. Origine et Preuve de Concept

L’auteur a navigué dans des conditions de post-pénurie depuis sa naissance — sécurité financière présente dès le départ, à travers de multiples cultures et continents, sans la pression organisatrice de la nécessité de survie. La question de ce à quoi sert une vie humaine, lorsque l’abondance matérielle est le point de départ, a été le terrain permanent de cette vie. C’est précisément la condition que l’AGI est prédite de produire à l’échelle civilisationnelle. Le schéma de 2013 a précédé le discours actuel d’une décennie.

Conclusion

> *Un système qui ne peut satisfaire la norme de générativité constitutionnelle n’est pas aligné. C’est de l’extraction portant le masque du progrès.*

Cette norme a été articulée pour la première fois en 2013. Le monde est arrivé à la question dont elle a toujours été la réponse.

Première articulation : schéma du bénéficiaire, mémoire EBS Paris, K. Berger, décembre 2013.
Développé sous sa forme complète par co-création humain-IA, mars 2026.