

Technologies de paramétrage algorithmique de l'expression émotionnelle et collaboration numérique : enjeux techno-moraux de la confiance interpersonnelle

Candidate : Nadia Guerouaou • **Encadrement :** Lionel Obadia et Olivier Alexandre • **Laboratoire d'accueil :** Centre Internet et Société (CNRS UPR 2000 / GDR 2091)

Contexte et problématique

En été 2024, *Soft Bank* géant des télécommunications japonais annonçait la sortie en 2026 d'*Emotion Cancelling*, son dispositif à destination des call-centers permettant de gommer la colère des clients mécontents. En Europe, la société *Alta Vocce* propose déjà d'appliquer de faux sourires sur les voix des appelants via *Alta Call* un équipement implémentable sur n'importe quelle plateforme de visio-conférence. Ces exemples posent la question du devenir de valeurs essentielles à la collaboration numérique comme la confiance lorsque notre expressivité émotionnelle peut s'affranchir de ses conditions biologiques et sociales et être artificiellement paramétrée par des technologies algorithmiques - médiatrices de nos interactions en ligne. L'importance de la dimension émotionnelle dans la collaboration interindividuelle est bien documentée. Les études analysant des échanges de binômes engagés dans des tâches collaboratives indiquent un bénéfice de l'expression des émotions pour l'augmentation de la créativité [10], la résolution de conflits et l'apprentissage collaboratif [3].

Du côté de l'Interaction Humain-Machine (IHM), l'importance de la "présence émotionnelle" se manifeste dans les nombreux travaux en informatique émotionnelle visant à conférer une expressivité émotionnelle à des agents artificiels pour optimiser la collaboration [12, 6]. Depuis récemment, c'est l'expressivité humaine qui est paramétrable et certaines études ont montré un effet positif de l'affichage d'un sourire artificiel pour la collaboration numérique. A l'instar de la créativité en télétravail collaboratif, augmentée par l'affichage de sourires artificiels chez les collaborateurs au moyen de *Smart Face*, un système de chat vidéo permettant de modifier les expressions faciales en temps réel lors des visioconférences [5]. À noter que ces transformations sont réalisées à l'insu des individus.

Avec l'augmentation post-pandémie de nos collaborations en ligne, ces technologies implémentables dans divers outils de visio-conférence sont au seuil de notre société. Dès lors, elles soulèvent une question de taille : au-delà de leurs potentiels effets à court-terme sur la performance collaborative, comment ces technologies de travestissement du réel pourraient-elles reconfigurer les valeurs qui sous-tendent la collaboration comme la confiance interpersonnelle ? Cette question relève des *changements techno-moraux*, à savoir les effets indirects mais structurels que les technologies peuvent avoir sur nos normes, nos valeurs et nos seuils d'acceptabilité [9]. Les technologies de collaboration en tant qu'artefacts socio-techniques ne sont pas neutres mais imprégnées de valeurs, dont celles de leurs concepteurs, qui interagissent avec celles des utilisateurs par la médiation technologique [11]. Cet enjeu est d'autant plus important que, compte tenu de la séparation usuelle (mais scientifiquement contestable) entre émotions et raison, ces systèmes pourraient être orientés vers la suppression des émotions dites "négatives" dans une perspective techno-solutionniste. Et influencer également nos considérations morales à l'égard de ce type d'émotions - avec des effets imprévus à plus long terme sur la collaboration numérique. Pourtant, en dépit du caractère évident de manipulation du réel de ces dispositifs, aucune étude ne s'est encore penchée sur leurs effets au regard de la confiance interindividuelle - dont le rôle médiateur sur la collaboration numérique est pourtant bien documenté [7]. La question de la confiance face aux manipulations algorithmiques du réel demeure cantonnée aux dispositifs à visée explicitement malveillante, tels que les *deepfakes* [8]. Les outils de collaboration émotionnellement augmentée, présentés comme bienveillants échappent largement à cet examen critique alors même que de tels effets pourraient entraver la collaboration numérique.

Objectifs et questions de recherche

Bien que le paramétrage algorithmique des émotions semble avoir des effets positifs à court terme sur des tâches collaboratives, il pose la question des changements techno-moraux opérant à plus long terme. En particulier sur la confiance inter-individuelle, et les effets retours sur la collaboration. En nous intéressant spécifiquement à la collaboration numérique lors d'une tâche de *délibération collaborative*, nous tâcherons de répondre à quatre questions de recherche :

QR1 : Les technologies de paramétrage émotionnel influencent-elles la confiance interpersonnelle médiatrice d'une collaboration numérique efficace ? Et dans quel sens ?

QR2 : La possibilité de contrôler l'expressivité émotionnelle pour optimiser la collaboration numérique modifie-t-elle notre acceptabilité de certaines émotions ? Ainsi dans une précédente étude interrogeant l'acceptabilité morale de divers scénarios d'usages de filtres vocaux émotionnels, nous avons mis en évidence que les usages

visant à diminuer la colère ou la nervosité dans la voix des utilisateurs étaient encore mieux acceptés que ceux permettant de créer un sourire artificiel [2].

QR3 : Quel est le référentiel de valeurs des concepteurs de plateformes d'optimisation de la collaboration numérique via le paramétrage des émotions concernant les QR 1 et 2.

QR4 : Comment les deux niveaux de changements techno-moraux des QR1 et 2 peuvent-ils impacter la collaboration numérique à moyen-long terme ?

Fondements théoriques

Théorie de la médiation morale technologique et changements techno-moraux

L'approche des *changements technomoraux* vise à *anticiper* l'influence des technologies sur les cadres éthiques et sociaux, afin d'inspirer à la fois les pratiques de conception des technologies et l'élaboration des politiques publiques. Il s'agit d'une méthode structurée d'anticipation des impacts indirects, fondée sur des recherches empiriques et sur l'analyse des pratiques existantes susceptibles d'être modifiées par l'introduction de nouvelles technologies [11]. Si l'on s'intéresse plus aisément à la moralité des individus à l'égard des technologies - ce fut l'objet de mon travail sur l'acceptabilité des technologies de transformation de l'émotion dans la voix [2] - il nous faut également interroger l'effet des valeurs charriées par ces technologies (dont celles de leurs concepteurs) sur la moralité des individus.

Confiance interpersonnelle et collaboration numérique

La confiance interpersonnelle constitue le mécanisme fondamental permettant la collaboration : elle régule la prise de risque sociale nécessaire à toute action collective [7]. La littérature distingue deux dimensions interdépendantes de la confiance [4] : (1) la confiance cognitive, fondée sur l'évaluation de la compétence et de la fiabilité du partenaire, et (2) la confiance affective, fondée sur les liens émotionnels et la perception d'un soin authentique mutuel. Le paramétrage algorithmique des émotions adresse spécifiquement la confiance affective et pourrait ainsi compromettre le processus collaboratif.

Neurosciences constructivistes et technologies collaboratives

À partir des modèles récents des neurosciences constructivistes (cerveau prédictif et émotions construites) développés dans ma thèse au sujet de l'influence des filtres vocaux émotionnels, nous posons la porosité fondamentale entre nos processus socio-affectifs indispensables à la collaboration et ce dispositif de collaboration numérique. Cette perspective permet d'analyser comment une modification algorithmique de l'expression émotionnelle peut transformer les processus de construction de la confiance et, par extension, la dynamique collaborative (régulation socio-émotionnelle, construction du consensus, confiance interpersonnelle).

Encadrement inter-disciplinaire

Approche socio-anthropologique

La co-direction avec Olivier Alexandre (CR CNRS, CIS, sociologue spécialiste de la Silicon Valley, de la culture et du numérique) et avec Lionel Obadia (PR Université Lyon 2, LARHRA UMR 5190, anthropologue des transformations culturelles associés aux technologies), permettra de compléter notre grille d'analyse théorique.

Approche méthodologique

1. **Revue systématique confiance et collaboration et analyse conceptuelle (2 mois)**
2. **Études empiriques (8 mois)**

Étude 1 - Dispositif expérimental test : (conception, réalisation et analyse des données :) Notre tâche constitue une situation de délibération collaborative entre deux partenaires sur des sujets polarisants (climat, IA à l'école) pour co-construire une position commune acceptable pour les deux parties. La production d'un document commun matérialisant le résultat du processus de délibération collaborative. Les participants (binômes) réaliseront la tâche de délibération collaborative via la plateforme de visioconférence DuckSoup lors de 3 phases. *Phase 1 - Condition contrôle (baseline)* : Tâche collaborative sans manipulation algorithmique de l'expression émotionnelle. *Phase 2 - Condition manipulation avec transparence variable* : Tâche collaborative avec manipulation algorithmique active (augmentation du sourire, atténuation des expressions de colère ou nervosité). Les participants sont assignés aléatoirement à deux sous-conditions :
• Groupe « Informé préalablement » : Les participants sont avertis avant la tâche de possibles manipulations de l'expression émotionnelle par leur binôme. • Groupe « Informé a posteriori » : Les participants sont

informés de la manipulation uniquement après la tâche. *Phase 3 - Condition incertitude - manipulation de la confiance inter-personnelle* : Tâche collaborative sans manipulation algorithmique réelle, mais tous les participants sont informés en début de phase que leur interlocuteur aura la possibilité de manipuler son expression émotionnelle. Cette condition teste l'effet de la simple croyance en une possible manipulation sur la confiance interpersonnelle.

Diverses mesures seront collectées à chaque phase : • De la collaboration : par codage des transcriptions selon un schéma validé [1] capturant les interactions socio-émotionnelles et les marqueurs de consensus. • Des ressentis perçus par les binômes : Échelles de Likert évaluant : (a) l'accord avec le texte final produit, (b) la perception du consensus pendant la délibération, (c) le degré de convergence perçu vers une position commune. Échelles de confiance interpersonnelle pré/post-tâche. *Mesures qualitatives (collectées à chaque phase)* : • Entretiens semi-directifs post-tâche explorant : (a) l'expérience subjective de collaboration, (b) l'évolution de la confiance avec leur partenaire, (c) l'acceptabilité morale des émotions perçues chez l'interlocuteur.

Étude 2 - Enquêtes auprès des concepteurs : Entretiens semi-directifs avec les développeurs de plateformes collaboratives intégrant des technologies de paramétrage émotionnel : *Alta Vocce*, *DuckSoup*, équipe de Shigeo Yoshida du *Smart Face* au Japon - avec lesquels nous avons collaboré au cours de notre travail de thèse. Ces entretiens exploreront les intentions de design, les présupposés normatifs, et les retours d'usage observés notamment pour la start up *Alta Vocce*, permettant d'analyser les scripts techno-moraux inscrits dans ces systèmes.

3. Construction de la grille des changements techno-moraux (2 mois)

Développement d'une grille d'analyse structurée des changements techno-moraux, synthétisant les résultats empiriques et les cadres théoriques mobilisés.

Résultats attendus et contributions

Contribution empirique

Des données originales sur l'impact des technologies d'augmentation émotionnelle sur la dynamique collaborative, combinant mesures quantitatives de la performance collaborative et analyses quantitatives et qualitatives de leur influence sur la confiance et l'acceptabilité émotionnelle. Les données visent à éclairer les choix de conception éthique des plateformes collaboratives intégrant du paramétrage émotionnel, en identifiant les risques de dégradation des dynamiques collaboratives.

Contribution méthodologique, éthique et dissémination

La grille techno-morale qui conformément à la méthodologie des changements techno-moraux permettra de poser les implications éthiques à trois niveaux : Micro (individuel) : transformation des critères de jugement de la confiance émotionnelle et des attentes expressives envers les interlocuteurs, Méso (relationnel) : normalisation de l'"optimisation émotionnelle" dans les interactions médiatisées, émergence de nouvelles normes d'expression ("toujours positif"), pression sociale à l'adoption ? Macro (sociétal) : érosion ou redéfinition de la confiance interpersonnelle comme ressource sociale, reconfiguration des normes d'expression émotionnelle légitime dans la collaboration numérique ? Notre protocole expérimental contribue à définir comment une innovation présentée comme "optimisation de la collaboration" transformerait le rapport à la confiance et à l'expressivité émotionnelle et les effets retour sur la collaboration.

Production d'articles scientifiques, de communications orales, et organisation d'un workshop interdisciplinaire réunissant chercheurs en STS, IHM et philosophie de la technique.

Positionnement dans le PEPR eNSEMBLE

Ce projet s'inscrit naturellement dans **PC 5 (TRANSVERSE)**, particulièrement le *Thème 1* en développant des méthodes pour quantifier et analyser l'impact des technologies de paramétrage émotionnel sur l'efficacité de la collaboration, et le *Thème 3* en traitant une question éthique importante : celle de la constitution de la moralité individuelle et sociétale à travers les dispositifs de collaboration numérique augmentés dans un contexte où les émotions sont souvent identifiées comme vecteurs de polarisation. L'échelle dyadique de notre protocole permet d'observer les micro-mécanismes qui, par l'adoption généralisée de ces technologies, transformeraient dans un second temps les fondements sociétaux de la confiance et de la collaboration : chaque interaction algorithmiquement médiée participe à la normalisation collective de nouvelles attentes relationnelles.

Le projet contribue également à **PC 1 (CATS)**, en étudiant la collaboration synchrone dans des espaces numériques médiatisés et en proposant des mesures pour évaluer la collaboration augmentée en environnements immersifs ou transformés algorithmiquement.

Références

- [1] Lei Gao, Xiaoran Li, Yaqian Zheng, Yanyan Li, Deliang Wang, and Haogang Bao. Discourse moves and socio-emotional interactions in collaborative argumentation : Exploring how their co-occurrence impacts argumentation quality. *Thinking Skills and Creativity*, 52 :101525, 2024.
- [2] Nadia Guerouaou, Guillaume Vaiva, and Jean-Julien Aucouturier. The shallow of your smile : the ethics of expressive vocal deep-fakes. *Philosophical Transactions of the Royal Society B*, 377(1841) :20210083, 2022.
- [3] Xiaoshan Huang and Susanne P. Lajoie. Social emotional interaction in collaborative learning : Why it matters and how can we measure it ? *Social Sciences & Humanities Open*, 7(1) :100447, 2023.
- [4] Daniel J. McAllister. Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1) :24–59, 1995.
- [5] Noboru Nakazato, Shigeo Yoshida, Sho Sakurai, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. Smart Face : enhancing creativity during video conferences using real-time facial deformation. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 75–83. ACM, 2014.
- [6] Jekaterina Novikova, Leon Watts, and Tetsunari Inamura. Emotionally expressive robot behavior improves human-robot collaboration. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 7–12. IEEE, 2015.
- [7] Laura M. L. Peters. *Now you see them... now you don't : Toward a greater understanding of virtual team effectiveness*. PhD thesis, University of Massachusetts Amherst, 2003.
- [8] Cristian Vaccari and Andrew Chadwick. Deepfakes and disinformation : Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social media+ society*, 6(1) :2056305120903408, 2020.
- [9] Simone van der Burg. Taking the “soft impacts” of technology into account : broadening the discourse in research practice. *Social Epistemology*, 23(3-4) :301–316, 2009.
- [10] Eva Vass. Exploring processes of collaborative creativity—The role of emotions in children’s joint creative writing. *Thinking Skills and Creativity*, 2(2) :107–117, 2007.
- [11] Peter-Paul Verbeek. Toward a theory of technological mediation : A program for postphenomenological research. In *Technoscience and postphenomenology : The Manhattan papers*, page 189. Lexington Books, 2016.
- [12] S. Zhou and L. Tian. Would you help a sad robot ? Influence of robots’ emotional expressions on human-multi-robot collaboration. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1243–1250. IEEE, 2020.