L’éthique en situations : une recherche-intervention avec les concepteurs d’une IA émotionnelle pour la famille

Ethics in situations: an action research with the designers of an Emotional AI for Family

< Julien PIERRE [1] > < Marie-Julie CATOIR-BRISSON [2] >

1. Département de communication, Université de Sherbrooke
julien.pierre@usherbrooke.ca

2. Département Communication, Culture et Langues, Audencia Nantes
mjcatoirbrisson@audencia.com

DOI : 10.25965/interfaces-numeriques.5202

< RÉSUMÉ >

Cet article rend compte d’une recherche-action conduite avec les concepteurs d’une IA émotionnelle qui tentaient de s’éloigner du modèle de capture des technologies affectives. Un tel déplacement implique de rattacher une approche critique et des pratiques d’intervention. Saisir et faire agir forment un couple qui vient compléter une éthique des devoirs ou des conséquences, non plus en aidant les concepteurs à arbitrer entre des principes ou des calculs, mais en leur offrant de repartir de leurs expériences, et celles de leurs usagers, pour concevoir une innovation socialement désirable. Cette promesse est rendue possible par une éthique en situations, un cadrage opérationnel permettant de saisir, par la voie d’une ethnographie sensible, ce qui est à disposition dans les situations vécues par les usagers et par les concepteurs ; de mettre à leur disposition, par des approches participatives, des objets intermédiaires facilitant la transition vers des imaginaires autres ; et de se tenir à disposition pour des médiations nouvelles, en s’appuyant sur des méthodes créatives et projectives.

< MOTS-CLÉS >

co-design, éthique, intelligence artificielle, activités de conception, méthodologie

< ABSTRACT >

This article discusses an action research project conducted with the designers of an emotional AI who were trying to avoid the capture model of affective technologies. This shift involves linking a critical approach to intervention practices. Seizing and acting constitute a couple that completes an ethics of duties or consequences, no longer by helping designers to arbitrate between principles or calculations, but by offering them to start again from their experiences, and those of their users, to design a socially desirable innovation. This promise is supported by a situational ethic, an operational framework that allows us to seize, through a sensitive ethnography, what is available in the situations experienced by users and designers ; to make available to them, through participatory approaches, intermediate objects that facilitate the transition towards alternative imaginaries ; and to be available for new mediations, based on creative and projective methods.

< KEYWORDS >

co-design, ethics, artificial intelligence, design phase, methodology

---

« Refaites vos choix »,
Énée, dans L'éveil d'Endymion,
Dan Simmons, 1998

Introduction

L'appel à la responsabilité est souvent réduit à la question de l'imputabilité : qui doit répondre, a posteriori, des effets de bord des technologies ? Or la responsabilité peut aussi s'entendre comme une manière de réduire, a priori, ces effets (Hache, 2011). Il est alors possible de se demander « comment répondre », par exemple aux demandes de développement d'une IA « responsable ». Dans cet article, nous revenons sur la manière dont nous avons accompagné une équipe de concepteurs qui se demandait comment répondre : d'un côté, à une demande d'innovation qu'ils identifiaient comme potentiellement nuisible ; et de l'autre, à un public espérant que leurs usages ne produisent pas d'effets négatifs a posteriori. Plus globalement, comment répondre à une planète dont les ressources seraient encore plus exploitées par une telle innovation ?

Le caractère systémique de ces enjeux appelle à dépasser le prisme exclusif de l'ingénierie, logicielle ou juridique, pour privilégier des

approches interdisciplinaires. À ce titre, les SHS peuvent contribuer au développement responsable des IA en fournissant de nouveaux jeux de données, en exhumant les ancrages socio-historiques des technologies, et en éclairant leurs effets sociaux (Velkovska & Relieu, 2021). Il reste cependant à transférer les connaissances ainsi produites dans les activités de conception, afin de soutenir celles-ci à partir des situations que l'innovation va transformer (Fischer, 2019). Un tel soutien recouvre d'abord, dans une perspective critique, une somme de médiations à saisir entre acteurs industriels, régulateurs, usagers, et objets techniques. Mais aussi, dans une perspective transformative, une somme de médiations à produire par les méthodes de la recherche-intervention (de La Broise et al., 2022).

Pour contextualiser l'expérience de recherche, nous avons travaillé avec des innovateurs mandatés pour tester un concept industriel, en l'occurrence un service domotique basé sur une IA émotionnelle (1ère partie). De janvier 2020 à juin 2022, nous avons cherché à résoudre leur dilemme éthique. En effet, deux perspectives se dessinent quand il s'agit d'automatisation et d'émotions : une logique de capture (Agre, 1994) et une logique d'encapacitation (Peugeot, 2015). C'est un choix à priori binaire, mais il est facile de cacher la capture derrière un discours d'encapacitation, ou de transformer un tel souhait en outil de capture. C'est donc à la définition de ces seuils que nous avons travaillé : l'équipe pressentait la facilité de fournir un service de capture, aurait voulu privilégier un modèle d'encapacitation, mais ne se sentait pas outillée pour aller du premier au second. Pour répondre à ce type de dilemme, l'éthique normative oscille entre une éthique des devoirs – répondre à sa hiérarchie – et une éthique des conséquences – ne pas nuire aux publics – (2ème partie). Pourtant une troisième voie est possible, relevant de l'éthique des vertus, capable de répondre aux enjeux de responsabilité à condition de l'opérationnaliser. Nous présentons ainsi une « éthique en situations » et son application dans un cas concret, que nous précisons en termes de méthodes et de résultats (3ème partie).

1. Les IA émotionnelles

Le projet industriel vise à interroger l'opportunité de développer une IA ayant des capacités émotionnelles. Cette orientation est à la croisée des

technologies affectives (TA) et du paradigme de l'IA forte : même s'il s'agit pour l'heure d'un sujet d'exploration, plusieurs cas d'usage sont déjà envisagés. Dans le cas spécifique d'une IA émotionnelle (IAé) installée dans le foyer familial, ce type de projet soulève de nombreux enjeux, que nous abordons dans les parties suivantes.

1.1. Les technologies affectives

Les TA combinent des modèles théoriques issus de la psychologie des émotions et des neurosciences (principalement issus du modèle des émotions basiques de Paul Ekman) avec les progrès de l'informatique affective (initiée par Rosalind Picard) : dit simplement, les expressions émotionnelles peuvent être encodées/décodées de manière informatique, rendant possible la reconnaissance automatique des expressions émotionnelles et sentimentales, et la simulation de traits émotionnels par des artefacts logiciels ou robotiques. Ce faisant, les TA délaissent les autres manifestations du spectre affectif (humeur, tempérament, etc.) ainsi que les variables culturelles (comme la structure sociale des émotions chez R. Williams), contextuelles (le travail émotionnel chez A. Hochschild) et situationnelles (les affordances chez Gibson). Les TA ont été renforcées par le design des interactions humain-machine et l'addition de couches d'interface multimodale, souvent de nature anthropomorphique, pour faciliter l'expérience-utilisateur.

Récemment, les capacités des TA ont été augmentées par la puissance de calcul combinatoire des réseaux de neurones artificiels. L'apprentissage-machine permet d'identifier des schémas émotionnels récurrents, dans les stimulations, les expressions, et la performance des recommandations. L'entraînement de ces modèles n'échappe pas à la reproduction de biais dans la détection de l'âge, du genre ou des origines ethniques, perpétuant les discriminations envers des populations déjà minorisées (McStay, 2018). À l'heure des IA génératives, le débat porte sur l'alignement moral des réponses faites par ChatGPT et consorts.

Malgré ces limitations, nous voyons dans ces technologies plusieurs intérêts de recherche, dont le premier est de convoquer une pensée du corps et des émotions en regard des systèmes techniques qui nous environnent : en soulignant la spécificité humaine de l'activité

émotionnelle face à des logiciels hyper rationalistes ; en resituant ces rapports avec la dimension praxique de nos usages ; en révélant aussi nos attachements à ce type d'artefacts.

1.2. Cas d'usage

Il nous faut ici faire un détour par les logiques d'innovation pour saisir les enjeux présentés dans la prochaine section. Deux approches nous semblent complémentaires, en termes de marché – ou marketing de l'innovation, et de processus – ou management de l'innovation.

Dès l'origine, Rosalind Picard prévoyait que l'informatique affective pourrait s'avérer utile dans le soin comme dans le divertissement (Picard, 2000). Les premiers cas d'usage sont avérés dans des marchés de niche (soutien thérapeutique, sécurité routière). Tandis que la logique des marchés de niche tend à optimiser l'efficience technologique, celle des marchés de masse tend à optimiser la rentabilité du produit. Afin d'assurer un retour financier, les investisseurs identifient de nouveaux marchés, quitte à travestir les finalités premières de la technologie. Ces technologies d'écriture/lecture automatisées des émotions sont dorénavant commercialisées dans les franges de l'économie de l'expérience et de la recommandation : divertissement, commerce, travail, pédagogie, etc. Les fonctionnalités proposées, en prise directe avec les émotions des usagers, tendent à reproduire, voire renforcer une forme de travail émotionnel et affectif (Alloing & Pierre, 2021 ; Negri & Hardt, 1999).

En termes de processus, le questionnement du partenaire industriel relève de l'innovation d'exploration. À la différence de l'innovation d'exploitation qui commercialise un produit, elle est dédiée à éprouver des idées, puis des concepts. Cette démarche est triple : il s'agit d'en sonder la faisabilité technique, la viabilité économique, et l'acceptabilité sociale. Notre mandat de recherche portait sur ce dernier point, en vue de tester sous quelles conditions l'idée d'une IAé mise à disposition des familles pouvait devenir un concept industriel.

L'autre intérêt de cette recherche se situe dans ce moment pivot de l'innovation, en sortie de laboratoire, au croisement de considérations

techno-scientifiques et financières, et qui décide in fine de ce qui est mis en société. À la manière de la philosophie expérimentale – qu'on pense par exemple au dilemme du tramway (Ogien, 2011) –, l'innovation d'exploration appelle des questions et des méthodes radicales pour y répondre. Cette radicalité est la seule manière d'appréhender les risques et de provoquer des controverses, préalables à la concertation et à la définition partagée d'un cadre éthique. Nous avons donc élaboré le scénario radical où une IAé pourrait être considérée comme un membre à part entière d'une famille, avec pour contrainte créative que cette IA ne reproduise pas des modèles identifiés comme nuisibles.

1.3. Enjeux

Une partie de la recherche consistait en l'élaboration d'une revue de littérature sur « la circulation des émotions dans la maison connectée ». Le premier objectif était de fournir un vocabulaire notionnel commun à une équipe pluridisciplinaire, et le second de proposer des éléments conceptuels apportant une assise critique au travail à venir. L'enjeu était de décentrer leur approche en présentant de nouveaux points de fuite socio-économique et anthropologique. Pour cela, nous avons montré dans quelles filiations s'inscrivait ce type de technologie. En effet, un modèle hégémonique se dessine derrière les technologies affectives, qui tend à imposer aux sociétés des trajectoires sans qu'elles fassent l'objet d'un consensus explicite, comme ce fût le cas lors de l'établissement du modèle thermo-industriel dont on mesure pleinement la nuisance aujourd'hui (Gras, 2013). Concernant les TA, nous repérons trois axes stratégiques dans cette avancée hégémonique : l'omniprésence du modèle ekmanien des émotions dans tous les corps de métier participant à la conception et l'exploitation des TA (Alloing & Pierre, 2017) ; l'inscription de ce modèle dans celui plus large de l'extractivisme, combinant l'extraction de ressources naturelles et celle des données personnelles (Chagnon et al., 2021) ; et la mise au service de ces logiques dans une forme de colonialisme (Couldry & Mejias, 2019). Ces logiques invasives et injonctives pénètrent les corps et les espaces privés, ici la maison et la famille, par le travail affectif requis et l'assignation d'affects aux individus (Hillis et al., 2015). Une même tendance hégémonique se dessine dans la conception de l'IA, avec un intérêt exclusif pour le modèle

connexionniste et plus récemment les larges modèles de langage, et une même exclusivité pour la déontologie dans la régulation.

Pour permettre aux concepteurs d'échapper à ces modèles, nous avons pris le parti de délaisser la technologie pour porter la réflexion sur le public visé : la famille. D'abord en la considérant comme un collectif, à contrario des tendances à la plateformisation du foyer (Goulden, 2019) qui tend à individualiser les pratiques, et les relations entre les usagers. C'est là aussi un modèle hégémonique dont il a fallu se défaire, issu du design centré utilisateur et des approches internalistes de la psychologie des émotions. En réponse, nous avons choisi des approches relevant de la pragmatique. Une première façon de concevoir la famille est de la considérer comme instituée par la filiation. Une autre est de la considérer comme constituée par des pratiques ordinaires (les routines) ou extraordinaires (les ritualités). C'est donc au « faire famille » (Morgan, 2011) que pourrait participer une IAé. Développer anthropologiquement ce qu'impliquerait une telle participation nous rapproche du modèle de compagnonnage évoqué par Donna Haraway. Ainsi, la contribution réciproque entre espèces (Haraway, 2003), de l'ADN à l'habitat partagé, pourrait dépasser le vivant pour concerner aussi nos relations avec des artefacts, dont des IA.

2. L'éthique : entre tentations et tentatives

Depuis les progrès faits autour des réseaux de neurones artificiels, les éthiciens sont régulièrement convoqués. Ils ont à leur disposition un outil à priori robuste : l'éthique normative, un appareillage de pensée permettant d'établir des normes comportementales en regard de valeurs ou principes moraux et servant de guide à la résolution des dilemmes éthiques. La tentation est grande de puiser dans cet appareillage pour encadrer le travail de conception et réguler les usages de l'intelligence artificielle. Nous présentons ce que ces tentatives ont donné, puis celle que nous avons mise en place.

2.1. Les deux facettes de l'éthique normative

Depuis Aristote, l'éthique se définit comme une réflexion sur l'action afin d'atteindre « quelque bien ». En mettant les conséquences en perspective, cette approche téléologique se double d'une logique évaluative, faisant la balance entre les moyens engagés et les fins obtenues. Dans cette éthique, la responsabilité (« qui doit répondre ? ») est mise en exergue. À l'inverse, une approche impérative, inspirée de Kant, évalue l'action en fonction de devoirs à respecter. C'est le régulateur ayant édicté les devoirs qui endosse la responsabilité. Il faut voir ces deux approches comme les bornes d'un continuum. À l'épreuve de leur application, notamment dans le cas des IA, ces approches peuvent se rejoindre. Nous verrons toutefois que cette convergence se traduit par une forme de solutionnisme technique.

En ce qui concerne les IA, la réflexion éthique est centrée sur l'algorithme, et sur ceux qui le conçoivent. Cette réflexion provient de Luciano Floridi et son éthique de l'information, qui permettrait de réduire l'entropie générée par l'infosphère (Floridi, 2013). L'auteur s'appuie sur l'approche normative d'Habermas pour espérer l'établissement d'un sens commun entre les parties prenantes. Ce souci de convergence empêche de discuter les biais inhérents au modèle hégémonique de l'IA connexionniste (Zacklad & Rouvroy, 2022). De plus, la focalisation de l'éthique sur les algorithmes délaisse le questionnement des interfaces, qui pourtant instancient leur présence en société. La seule interface est alors de papier, la réflexion éthique se réfugiant derrière la rédaction de principes, aboutissant à une forme de « saturation éthique » (Noiseau, 2023). Parmi toutes les chartes et codes déontologiques, L. Floridi et son équipe en identifient cinq principaux (Morley et al., 2020) : deux principes conséquentialistes (bienfaisance et non-malfaisance), et trois principes déontologiques (devoir de préservation de l'agentivité humaine, devoir de justice par effacement des biais, et devoir d'explicabilité). Ils proposent alors de plateformiser cette démarche, selon un modèle d'Ethics-as-a-Service (Morley et al., 2021), un dispositif informationnel tenant ces principes à disposition des développeurs, selon leurs besoins. Cette solution déléguée à la technique efface toute forme d'intermédiation humaine dans la concertation éthique, elle reproduit aussi une forme de « magasinage éthique », pourtant dénoncée par

Floridi. Enfin, les auteurs accordent à ces principes moraux un caractère paternaliste, effaçant la possibilité que les concepteurs se les réapproprient. Or, les multiples tentatives pour établir des principes éthiques se sont avérées inopérantes (Mittelstadt, 2019), parce que ces principes sont de fait en situation de « ré-énonciation permanente » au cours des activités de conception (Hoang et al., 2022 ; Mitropoulou & Pignier, 2018).

2.2. Une éthique en situation(s)

Rien ne dit en quoi obéir à des principes ou confier ses actions au calcul permet d'atteindre « quelque bien ». Au regard de ces éthiques, nous revendiquons un pas de côté et plaidons pour l'éthique des vertus, dernière branche de l'éthique normative. En guise de plaidoyer, c'est un retour d'expérience scientifique que nous proposons, précédé de l'élaboration conceptuelle faisant de « l'éthique en situations » la contribution première de cet article. Celle-ci prend source dans l'éthique des vertus, dans les approches situées, enfin dans les démarches processuelles.

Une première définition de l'éthique des vertus propose d'imiter des êtres vertueux (Ogien, 2011). Ce faisant, elle délègue la responsabilité au « héros », sans donner d'autre méthode que l'imitation. Une autre définition resitue la responsabilité chez l'individu, en l'invitant à orienter ses actions selon ses propres vertus (en connaissance de ses forces et faiblesses), mais aussi selon celles de ses interactants. C'est une invitation à apprendre à faire selon les dispositions : les siennes, celles des autres, et celles qui se tiennent dans l'environnement. Cette réflexion nourrit l'éthique du soin à destination des personnes vulnérables (Tronto, 2009) ; elle se rapproche des éthiques féministes appliquées au numérique (Mellot et al., 2022b), de la maintenance de nos infrastructures et de ceux qui les habitent (Hine, 2020) ; et se retrouve aussi dans le domaine des IA (Hagendorff, 2020 ; Noiseau, 2023 ; Toupin, 2023).

L'éthique des vertus rejoint l'éthique de la puissance chez Spinoza (1677), où les vertus sont des manières d'être affecté et d'affecter ce, celles et ceux qui nous entourent. En plus d'être une philosophie qui

explique, elle assume d'être une orientation pour penser ses actions. En effet, l'éthique de Spinoza permet de penser les actions (individuelles ou collectives) dans une dialectique entre un pouvoir qui capture et une puissance qui encapacite (Negri, 1982). Ainsi, le développement d'une IAé située au sein du foyer familial, avec l'intention de contribuer au « faire famille » ne peut se faire qu'en mobilisant une éthique de la puissance. Dans cette perspective, rendre désirable une IA revient à montrer dans quelles situations elle augmente la puissance d'agir, comment elle repose sur les forces à disposition sans produire, reproduire ou déplacer des faiblesses. Cela implique aussi de montrer quelles situations ont nourri sa conception.

Les connaissances sont situées dans des rapports de pouvoir, de même que l'éthique est située dans des tensions (Hache, 2011). L'éthique située permettrait ainsi de remettre en cause les présupposés scientifiques de l'automatisation algorithmique (Zacklad & Rouvroy, 2022), et ceux de la psychologie des émotions à l'œuvre dans les TA. L'enquête pragmatique permet de dévoiler ces tensions et d'aborder autrement les émotions, en montrant comment le triple sens des émotions (sensation, orientation, signification) se construit dans la situation (Dewey, 1895 ; Quéré, 2021), qu'elle se déroule dans les foyers, ou dans les lieux où travaillent les concepteurs.

Dans ces situations, le plan est toujours débordé, quand l'usager ne suit pas les prescriptions d'une interface (Suchman, 1987), quand il s'agit de maintenir les routines familiales, ou de gérer un projet en entreprise. Dans ce dernier cas, les activités de conception sont situées dans des tensions organisationnelles, des méthodes dont les concepteurs sont eux-mêmes des usagers. L'enquête se doit alors d'aller dans ces moments où émerge le sens, quitte à les provoquer. Il importe de prendre en compte plusieurs niveaux comme dans l'éthique créative des technologies (Catoir-Brisson, 2019), une approche qui embrasse les niveaux individuels, sociaux et environnementaux. Dans cette perspective, l'éthique est un processus en tension tout au long du projet de conception d'une technologie, qui porte les valeurs spécifiques de celles et ceux qui l'ont conçue (Fabris, 2018). S'ouvre alors une série de questions : quelles situations permettent de saisir la vie affective en famille, comment une

IAé pourrait s'en saisir, et quelles situations produire pour que les concepteurs s'en saisissent ?

Cette approche affective dévoile un maillage plus complexe que le modèle de l'encodage. Son dévoilement est processuel, abductif, tant les situations vécues appellent à réécrire le sens donné aux objets de l'enquête. À rebours des approches protocolaires de la déontologie, l'éthique en situations est un apprentissage continu basé sur plusieurs principes qui ont guidé et ont émergé de notre recherche : ils s'appliquent aussi bien au chercheur, au concepteur, et à l'usager. C'est d'abord une éthique qui apprend à saisir, par une forme de « bienveillance dispositive » (Belin, 1999), ce qui se joue dans les situations : la rencontre entre les dispositions affectives des individus et ce que les interfaces numériques disposent dans leur environnement (et comment elles disposent des traces que laissent ces derniers). C'est ensuite une éthique qui apprend à agir dans ces relations, en choisissant les connaissances, affects, intentions (et données personnelles) à tenir à disposition de ses partenaires. C'est une éthique qui apprend à se tenir à disposition des sollicitations, désirées ou non, à saisir les médiations, les mettre à disposition, et en proposer de nouvelles. Cette éthique est donc un engagement méthodologique : c'est à travers elle que nous avons élaboré nos questions de recherche, puis nos questions de méthode.

3. Mise à l'épreuve : la conception familiale d'une IA émotionnelle

Définir ce que serait une éthique en situations ne peut venir sans une mise à l'épreuve par des questions de méthode, initiées par les trois questions de recherche suivantes :

- Q1 (Saisir les médiations) : Où sont situées les connaissances dont ont besoin les concepteurs ? Et comment atteindre ces situations dans le respect de celles et ceux qui les vivent ?
- Q2 (Mettre ces médiations à disposition) : Dans quelle mesure la connaissance de ces situations permet-elle d'alimenter le travail des concepteurs dans le design de leur technologie ? Quels outils permettent de transférer les apports d'une observation des situations au sein du processus de conception ?

– Q3 (Tenir à disposition des médiations nouvelles) : Que produit cette encapacitation ? Et comment l'évaluer ?

3.1. Méthodologie

Les méthodes inspirées du modèle des émotions basiques reposent sur des capteurs biométriques utilisés lors de simulations en laboratoire, incompatibles avec une approche située. La démarche ethnographique s'avère plus adéquate pour approcher la vie affective au plus près. Les travaux relevant de l'ethnographie située (Velkovska & Relieu, 2020) ou de l'ethnographie affective (Gherardi, 2019) offrent des prises pour saisir cette intimité, sans reproduire le modèle invasif commun à une certaine ethnographie et à l'économie numérique des datas personnelles. Nous identifions trois leviers dans ces méthodes : une approche sensible (Pierre & Catoir-Brisson, 2023 ; Pink, 2015, 2021), participative (Catoir-Brisson, 2018 ; Grosjean, 2022 ; Martin-Juchat & Bonnet, 2023), et projective (Bleecker, 2009 ; Catoir-Brisson, 2019). Pour suivre la démarche abductive, nous présentons ci-dessous les différentes étapes de notre méthodologie, suivant les corpus que nous avons colligés tout au long de notre travail[1].

Considérant l'IA émotionnelle comme une machine empathique, nous avons proposé à nos participants, concepteurs et usagers, de prendre la place de la machine afin de mieux la construire. Plutôt que d'imposer notre propre outil de collecte (Martin-Juchat & Pierre, 2015), nous avons demandé à des étudiants volontaires comment ils pouvaient rendre compte de la vie affective de leur famille. Ils ont produit des documents de format divers (audio, vidéo, texte, tableau, dessins, photos) alignés sur le caractère multimodal des expressions émotionnelles. Les récits partagés faisaient état d'un travail affectif, diversement distribué selon les membres, les pièces, les moments. Sur cette base, nous avons développé un formulaire en ligne, dont l'interface – testée par tous les participants – a facilité la saisie du travail affectif. Avec cet outil, nous

1 Nous précisons aussi que le calendrier de notre projet se confond avec celui du confinement. Nous avons donc dû faire avec ce qui était à disposition : des maisons fermées aux (chercheurs) étrangers, des routines bouleversées, des émotions marquées par le stress ambiant.

nous sommes engagés dans une auto-ethnographie avec un triple objectif : montrer que les affects sont observables, que le dispositif n'est pas intrusif puisqu'il remplace la captation par la déclaration, et que cette observation peut fournir un jeu de données alternatif à une IA (Velkovska & Relieu, 2021). Ces données ont constitué le corpus de notre base documentaire, à disposition de nos prochains participants. De même que pour la collecte, nous avons laissé aux participants le soin de trouver du sens aux phénomènes que nous leur mettions à disposition. Nous avons poursuivi en leur demandant d'imaginer, puis d'éprouver, leur propre outil de saisie du travail affectif dans leur famille. Cette étape a permis d'insister sur les tensions propres au dialogue interprofessionnel en croisant des paradigmes de conception et des imaginaires différents. Notre démarche d'implication par le faire a aussi ouvert la voie à une réflexivité des concepteurs sur leur pratique de conception, à partir de leurs propres routines familiales. Cette démarche réflexive a permis de concevoir autrement, non pas à partir d'un usager incarné dans un persona fictif mais bien d'un collectif avec des habitudes ancrées dans l'espace domestique. Elle a aussi été nécessaire pour imaginer avec eux les dispositifs d'IAé qui correspondaient au mieux à leurs expériences quotidiennes. Nous avons accordé à l'ensemble de ces données (et aux outils qui les ont façonnés) le rôle d'une « ethnographie du présent ».

Notre deuxième corpus constitue une « ethnographie du futur ». Comme indiqué, l'IA forte n'existe pas de nos jours, encore moins l'IAé. Il n'y a pas matière à observer ces usages dans le foyer familial. Nous avons fait appel à l'imagination pour cela : celle de nos étudiants, des concepteurs, et des artistes. Un premier travail a consisté à repérer les sources d'inspiration des innovateurs. Dans le cadre de cours ou d'ateliers, nous avons demandé à des étudiants d'imaginer une IAé en s'appuyant sur celles qu'ils connaissaient. Nous avons posé la même question aux concepteurs, et nous avons réuni leurs réponses avec notre propre recension, en identifiant des IA de fiction, de marché, ou de laboratoire. L'ensemble de ces créatures constitue le Bestiaire des IA, un catalogue des vices et vertus des robots sociaux, assistants vocaux, et IA, réels ou fictionnels. Un dernier travail, considéré comme du design fiction, a visé l'écriture de « fabulations spéculatives » (Haraway, 2016) et la production de « prototypes diégétiques » (Sterling, 2009) permettant d'interroger le type de relation et d'interaction entre la

créature et les humains, non pas dans une relation d'usage mais de compagnonnage (l'un soutenant l'autre).

Enfin, un dernier corpus, intitulé « ethnographie de l'entre-deux », fait la passerelle entre le présent et le futur, et entre le domicile et l'entreprise. Il est constitué des travaux étudiants relatant les stratégies de communication préalables à une mise en société de l'IAé, et des productions des concepteurs relatant les stratégies de médiation avant la mise en production. Cette étape souligne les tensions qui surgissent quand vient le moment de produire des éléments de langage, entre des étudiants mandatés pour accompagner une entreprise (et insistant sur les enjeux de vie privée) ou pour imaginer des cas d'usage (insistant sur le caractère utopique/dystopique). Cette même tension se retrouve entre une équipe de conception et ce qu'elle perçoit des attentes de son équipe de direction. Dans cette perspective, nous avons conçu un dernier atelier, animé dans les locaux de notre partenaire, de manière à bien situer les enjeux organisationnels. Cet exercice a permis aux concepteurs de prototyper un service numérique à partir de toutes les connaissances et expériences mises à disposition, avec pour contrainte créative de présenter à leur direction le service sous forme de totem. La vie de l'organisation, faite de restructurations, n'a pas permis d'accomplir cette dernière étape. Cependant, le développement du Bestiaire pourrait permettre de compenser cette situation, en le configurant comme un outil d'aide à la décision[2].

Ainsi, nous répondons à la première question de recherche en investiguant des situations présentes (approche sensible de la vie affective des familles) et à venir (approche projective des fictions et design fiction) par des méthodes ethnographiques situées. En installant une démarche de co-conception des outils de collecte et de co-analyse des données ethnographiques (approche participative du co-design), nous favorisons un transfert de connaissances depuis la sphère des usages vers celle de la conception, répondant ainsi à la deuxième question de recherche. Enfin, nous allons voir dans la partie suivante comment nous évaluons la performance de ce travail.

2 Chapitre en cours de parution.

3.2. Résultats

Trois résultats majeurs ont émergé de ce travail : la formalisation d'une éthique en situation, un assemblage méthodologique capable d'opérationnaliser cette éthique, et par ces méthodes la production de concepts industriels répondant aux enjeux identifiés au préalable.

La production du Bestiaire (210 créatures à date) et des œuvres de design fiction (12 par les étudiants, 3 par les concepteurs) sont rassemblés dans un dispositif transmédiatique d'abord, médiationnel ensuite, rejoignant le format du « dispositif de transmédiation » (Zacklad & Catoir-Brisson, 2021). Un tel dispositif met à disposition des participants la recherche en train de se faire [3]. Le caractère transmédiatique tient au fait que les données sont distribuées dans plusieurs formats médiatiques, selon ce que la situation requiert. Par exemple, le Bestiaire est développé au format numérique (Ill. 1), mais une version du catalogue existe en cartes imprimées, telles que nous les avons utilisées en atelier (Ill. 2).

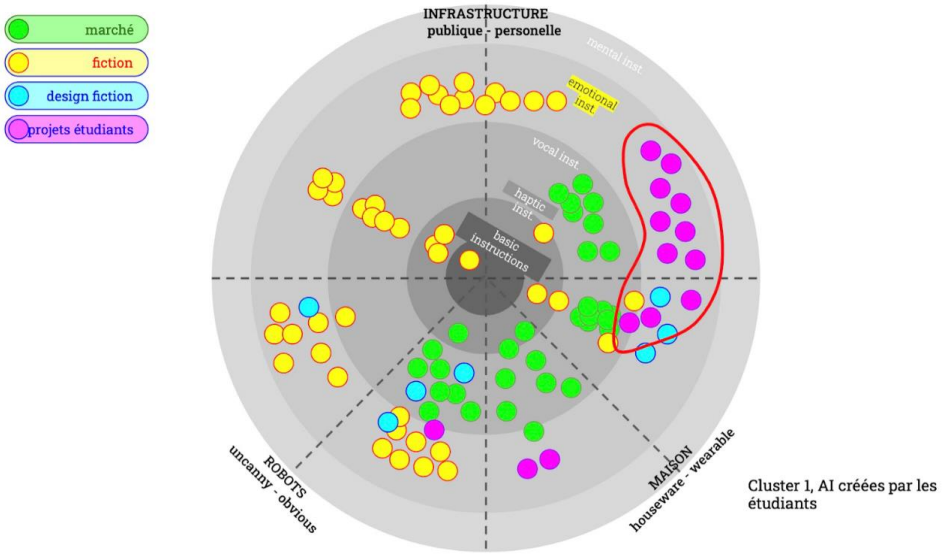

Illustration 1 : le bestiaire présenté en conférence sous forme de diagramme

3 Tous les résultats sont accessibles via les liens présentés en annexes.

Cette mise à disposition des contenus favorise le caractère médiationnel du dispositif. Ici, nous insistons sur la factitivité des objets intermédiaires (Bassereau et al., 2015 ; Vinck, 2009) conçus pendant l'enquête : de l'application auto-ethnographique au Bestiaire, des scénarios aux modèles de document favorisant la progression de la réflexion. En faisant faire, selon des méthodes créatives, narratives et visuelles (Mannay, 2015), en empruntant la méthode d'élicitation des photos symptômes (Catoir-Brisson, 2018) et en faisant appel à l'auto-ethnographie, les participants ont mis en discussion des considérations affectives, praxiques, et éthiques ouvrant de possibles nouvelles trajectoires contraires aux hégémonies actuelles.

Illustration 2 : le bestiaire sous forme de cartes à manipuler en atelier

À l'issue de ce processus, les concepteurs ont mis en forme trois services qui leur paraissent adéquats, ou à tout le moins qu'ils sont prêts à soutenir devant leur direction, et possiblement devant le public.

- Un dispositif de canalisation de l'énergie émotionnelle
- Un affichage des dispositions affectives des membres de la famille
- Un biographe familial pouvant rappeler les bons moments passés ensemble[4]

4 Ce projet nous semblant le plus prometteur, nous avons imaginé ses trajectoires possibles, entre conception et usage, dans un texte qui prendra la forme d'une recherche-fiction (à paraitre).

4. Discussion

Nous pouvons souligner le rôle des objets intermédiaires co-créés avec les participants et les mettre en lien avec le processus de l'éthique en situation. En effet, que ce soit les design fiction, les scénarios d'usage produits avec les étudiants, ou les prototypes de service co-créées avec les concepteurs, chaque média mobilisé a permis de produire des effets pour les faire réagir ou pour stimuler les imaginaires individuels et collectifs. Les études médiatiques ont montré comment la matérialité des objets intermédiaires pendant le processus de co-design permet d'ouvrir des perspectives nouvelles sur les trajectoires culturelles des technologies (Catoir-Brisson, 2022). C'est aussi le cas dans cette recherche : nous avons accompagné le passage chez les concepteurs de l'imaginaire d'une IAé forte à celui d'un objet à comportement. La contribution se situe aussi dans le fait que ces objets intermédiaires peuvent constituer des leviers de mobilisation par les parties prenantes, en circulant au sein de l'entreprise. L'éthique en situations est ainsi un processus en tension et la matérialité des objets intermédiaires mobilisés dépend des effets visés sur chaque destinataire (comme les cartes à manipuler en atelier, le diagramme du bestiaire pour le partager en recherche, les prototypes des dispositifs d'IAé pour les exposer dans le hall de l'entreprise, etc.).

Peu importe la pertinence de ce qui est produit, c'est le processus qui prévaut ici, et ce qu'il a permis de générer comme alternatives face aux doutes que l'équipe exprimait au départ. L'enjeu éthique est de rendre possible ces trajectoires alternatives et potentielles, et d'accompagner leurs actualisations et leurs mises en situation à partir des différentes formes de factitivité des objets intermédiaires (Catoir-Brisson, 2022) : des totems (faire faire) aux scénarios d'usage (faire savoir) ou aux prototypes de service (faire croire).

Ainsi, le modèle de captation des émotions a été abandonné au profit d'un modèle déclaratif ; la figure de l'IA forte a été écartée pour préférer une conception centrée sur des objets-à-comportement ; et un design centré collectif a été adopté. Au-delà de ces détails techniques, le parcours proposé a contribué à forger des considérations éthiques. Nous les présentons selon les six espaces de controverses propres à l'éthique du

numérique et de l'IA (Zacklad & Rouvroy, 2022) : dans l'espace de la culture et des croyances, les concepteurs ont remis en cause la figure d'une IA forte membre de la famille ; dans celui de l'emploi et des transformations sociétales a été actée la part de travail affectif que requièrent les TA en particulier, le numérique en général ; dans les controverses liées au travail et aux modalités d'organisation, la méthodologie a montré les apports des pratiques sensibles, participatives et projectives ; concernant les controverses sur la citoyenneté, la diversité et l'égalité des chances, ainsi que celles sur les enjeux écologiques, les discussions ont porté d'abord sur le caractère genré du travail affectif, dans la majorité des cas confié aux femmes, puis sur l'iniquité d'accès de telles innovations entre publics favorisés et publics vulnérables, enfin les considérations sur l'impact environnemental de l'IAé ont été intégrées dans les trois totems. Pour finir, la participation des publics aux choix techniques qui les impactent est un objet de controverse : notre contribution y est ici limitée dans la mesure où l'empan des méthodes participatives employées a été modeste, en termes de représentation sociale, et de volume de personnes ayant participé au projet. Nous envisageons, avec le Bestiaire comme dispositif de transmédiation, de déployer cette approche à un plus grand nombre.

5. Conclusion

Sans délaisser les éthiques des devoirs et des conséquences, l'éthique des situations est un complément utile à mobiliser quand la volonté se présente d'encadrer le travail de conception ou les usages des technologies émergentes. Cette utilité tient à son caractère opérant : si pour l'heure elle a été forgée sur un seul cas, elle dispose d'une flexibilité certaine pour s'adapter à d'autres situations. Cette opérationnalisation tient du tuilage proposé entre une recherche-critique permettant, dans une forme de maïeutique, de se saisir d'enjeux souvent invisibilisés parce que trop complexes ou hégémoniques, et une recherche-intervention qui, dans une forme d'heuristique, permet de faire émerger des alternatives pour échapper à des modèles jugés nuisibles. À ce titre, l'assemblage méthodologique que nous avons élaboré et présenté ici, fait d'approches sensibles, participatives, projectives, semble une contribution majeure à

l'heure où des formes nouvelles de collaboration sont requises pour résoudre des problèmes anciens.

Accompagner la conception (et l'abandon) d'une IAé a été un projet riche, conceptuellement et méthodologiquement. Cela a permis de répondre à la demande immédiate d'un industriel, mais aussi de développer un appareillage éthique original, qui reste à éprouver. Trois épreuves majeures sont à expérimenter, du point de vue de la recherche. Une épreuve de robustesse, auprès non plus des concepteurs mais cette fois des développeurs qui encodent (notamment ceux qui travaillent sur les IA génératives) ou des innovateurs qui œuvrent à améliorer « l'habitabilité du monde » (Findeli, 2015). Une épreuve de transfert, en portant cette démarche dans d'autres champs technologiques (l'informatique quantique par exemple). Une épreuve d'appropriation, en sondant comment l'éthique des situations peut permettre aux usagers de configurer leurs interfaces, et de choisir quel compagnonnage avec les IA leur paraît désirable.

Remerciements

Nous tenons ici à remercier Frédéric Massa et Laurence Dhaleine, respectivement chef de projet et sociologue au sein d'Orange Labs, pour la confiance accordée dès les premiers échanges, leur ouverture d'esprit à des propositions conceptuelles et méthodologiques, ainsi que leur soutien complet dans toutes les phases du projet. Nous souhaitons ici que cette contribution soit saluée.

Bibliographie

Alloing, C., & Pierre, J. (2017). Le web affectif : Une économie numérique des émotions. INA.

Alloing, C., & Pierre, J. (2021). Le travail émotionnel numérique : Faire de ses clics un moyen d'éviter les claques. Questions de communication, 40(2), 233-256. https://doi.org/10/gr5j38

Bassereau, J.-F., Charvet Pello, R., Faucheu, J., & Delafosse, D. (2015). Les objets intermédiaires de conception / design, instruments d'une recherche par le design. Sciences du Design, 2(2), 48-63. https://doi.org/10/ggws7d

Belin, E. (1999). De la bienveillance dispositive. Hermès, La Revue, n° 25(3), 243-259. https://doi.org/10/fcps75

Bleecker, J. (2009). Design Fiction: A short essay on design, science, fact and fiction. Near future laboratory, 29.

Catoir-Brisson, M-J. (2022). « La matérialité de la communication dans les approches de co-design : quelles contributions à la transformation dans les organisations ? » *ATIC* n° 4, 29-51.

Catoir-Brisson, M.-J. (2019). « Design social et science-fiction pour penser la ville intelligente au service de demain ». *13èmes Rencontre euro-méditerranéennes Volubilis*, 34-48.

Catoir-Brisson, M-J. (2018). « Social innovation by design in mobile healthcare for sleep disorders ». Proceedings of Design Research Society, Vol 1, p. 2324-2333.

Chagnon, C. W., Hagolani-Albov, S. E., & Hokkanen, S. (2021). Extractivism at your fingertips. Dans Our extractive age (p. 176-188). Routledge.

Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. Television & New Media, 20(4), 336-349. https://doi.org/10/gfj88j

Dewey, J. (1895). The theory of emotion. Psychological review, 2(1), 13. https://doi.org/10/fwfk87

Fabris, A. (2018). Ethics of information and communication technologies. Springer.

Findeli, A. (2015). La recherche-projet en design et la question de la question de recherche : Essai de clarification conceptuelle. Sciences du design, 1, 45-57. https://doi.org/10.3917/sdd.001.0045

Fischer, F. (2019). L'éthique by design du numérique : Généalogie d'un concept. Sciences du Design, 10(2), 61-67. https://doi.org/10.3917/sdd.010.0061

Floridi, L. (2013). The philosophy of information. OUP Oxford.

Gherardi, S. (2019). Theorizing affective ethnography for organization studies. Organization, 26(6), 741-760. https://doi.org/10/gf37rq

Goulden, M. (2019). 'Delete the family': Platform families and the colonisation of the smart home. Information, Communication & Society, 0(0), 1-18. https://doi.org/10/gjt7db

Gras, A. (2013). Les imaginaires de l'innovation technique : Regard anthropologique sur le passé dans la perspective d'un avenir incertain. Manucius.

Grosjean, S. (2022). Le co-design de technologies de eSanté : Un enchevêtrement de conversations, de tensions créatrices et d'inscriptions circulantes.

Approches Théoriques en Information-Communication (ATIC), 4(1), 103-125. https://doi.org/10.3917/atic.004.0103

Hache, É. (2011). Ce à quoi nous tenons. Propositions pour une écologie pragmatique. La Découverte.

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. Minds and Machines, 30(1), 99-120. https://doi.org/10.1007/s11023-020-09517-8

Haraway, D. J. (2003). The companion species manifesto: Dogs, people, and significant otherness (Vol. 1). Prickly Paradigm Press Chicago.

Haraway, D. J. (2016). Staying with the trouble: Making kin in the Chthulucene. Duke University Press.

Hillis, K., Paasonen, S., & Petit, M. (2015). Networked affect. MIT Press.

Hine, C. (2020). Strategies for Reflexive Ethnography in the Smart Home: Autoethnography of Silence and Emotion. Sociology, 54(1), 22-36. https://doi.org/10/ghmfgs

Hoang, A. N., Mellot, S., & Prodhomme, M. (2022). Le numérique questionné par l'éthique située des écologies politiques. Revue française des sciences de l'information et de la communication, 25, Article 25. https://doi.org/10.4000/rfsic.13239

Mannay, D. (2015). Visual, narrative and creative research methods: Application, reflection and ethics. Routledge.

Martin-Juchat, F., & Bonnet, F. (2023). De la communication au design : Instrumentalisation ou renouvellement ? Approches Théoriques en Information-Communication (ATIC), 6(1), 5-8. https://www.cairn.info/revue-approches-theoriques-en-information-communication-2023-1-page-5.htm

Martin-Juchat, F., & Pierre, J. (2015). Le numérique pour tromper l'ennui au travail : Usages affectifs des TIC par les jeunes adultes. http://hal.univ-grenoble-alpes.fr/hal-01374936

McStay, A. (2020). Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. Big Data & Society, 7. https://doi.org/10/ggsfrn

Mitropoulou, E., & Pignier, N. (2018). Le sens au cœur des dispositifs et des environnements. Connaissances & Savoirs.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501-507. https://doi.org/10.1038/s42256-019-0114-4

Morgan, D. (2011). Rethinking family practices. Springer.

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. Minds and Machines, 31(2), 239-256. https://doi.org/10.1007/s11023-021-09563-w

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. Science and Engineering Ethics, 26(4), 2141-2168. https://doi.org/10.1007/s11948-019-00165-5

Negri, T., & Hardt, M. (2004). Multitude : Guerre et démocratie à l'époque de l'Empire. Editions la Découverte.

Noiseau, P. (2023). Ethics of care and artificial intelligence: The need to integrate a feminist normative aproach. Dans B. Prud'homme, C. Régis, & G. Farnadi (Éds.), Missing Links in AI Governance (p. 349-362). UNESCO.

Ogien, R. (2011). L'influence de l'odeur des croissants chauds sur la bonté humaine : Et autres questions de philosophie morale expérimentale. Grasset.

Picard, R. W. (2000). Affective computing. MIT press.

Pierre, J., & Catoir-Brisson, M.-J. (2023, 05). Comment répondre à une IA émotionnelle ? Le choix d'une éthique située : Au plus près des usagers et des concepteurs. L'IA en pratique(s) : l'éthique est-elle automatique ? LabCMO + LabFluens + GENIC, Montréal, QC, CA.

Pink, S. (2015). Doing sensory ethnography. Sage.

Pink, S. (2021). Sensuous futures: Re-thinking the concept of trust in design anthropology. The Senses and Society, 16(2), 193-202. https://doi.org/10/gkmzqs

Quéré, L. (2021). La fabrique des émotions. Puf.

Sterling, B. (2009). Design fiction. Interactions, 16(3), 20-24. https://doi.org/10/cfx568

Suchman, L. A. (1987). Plans and situated actions: The problem of human-machine communication. Cambridge university press.

Toupin, S. (2023). Shaping feminist artificial intelligence. New Media & Society. https://doi.org/10.1177/14614448221150776

Tronto, J. (2009). Un monde vulnérable. Pour une politique du « care » (H. Maury, Trad.). Lectures, Les rééditions.

Velkovska, J., & Relieu, M. (2021). Pour une conception « située » de l'intelligence artificielle. Réseaux, 229(5), 215-229. https://doi.org/10/gn2w2k

Vinck, D. (2009). De l'objet intermédiaire à l'objet-frontière. Revue d'anthropologie des connaissances, 3(1), 51-72. https://doi.org/10/fkb99q

Zacklad, M., & Catoir-Brisson, M-J. (2021) « Culture de la conception et du design dans la recherche-intervention en SHS », RFSIC n° 23. https://doi.org/10.4000/rfsic.11860

Zacklad, M., & Rouvroy, A. (2022). L'éthique située de l'IA et ses controverses. Revue française des sciences de l'information et de la communication, 25, Article 25. https://doi.org/10.4000/rfsic.13204

Annexes

- Accès aux données de l'auto-ethnographie : URL : https://affect.wiki/famille/dashboard/emotions.php. Dernier accès : 04/01/2024
- Accès aux design fiction : URL : http://affect.wiki/designfiction/. Dernier accès : 04/01/2024
- Accès au Bestiaire : URL : http://bestiaireIA.net/. Dernier accès : 04/01/2024