

# Qui a mis des émotions dans la machine?

Nic Ulmi

L'empathie s'invite désormais jusque dans nos échanges avec les machines. Cette apparente bienveillance n'est pas le fruit du hasard, mais bien le résultat d'un travail délibéré. Le spécialiste de l'interface humain-machine Laurent Bolli analyse les enjeux de cette évolution.

«Je comprends. Je suis là pour toi. Tu veux en parler ? » dit ChatGPT si on lui confie : « Je suis triste ». Pendant une seconde, on y croit, on le ressent: l'IA a l'air de partager nos émotions. Et dans un sens, cette faculté affective est bien réelle. Simplement, elle n'appartient pas au chatbot vedette, mais aux équipes qui l'ont conçu. Car dans l'ingénierie du numérique, l'empathie se trouve au cœur du travail sur l'« expérience utilisateur » (lire l'article en p. 64 consacré à ce concept). Le professeur Laurent Bolli explore ce territoire dans ses cours sur l'«interface humain-machine» à la Haute École d'Ingénierie et de Gestion du Canton de Vaud - HEIG-VD - HES-SO.

Comment l'empathie apparaît-elle dans l'interaction avec une machine ? Une chose que je dis à mes étudiantes et étudiants, c'est que, dans un sens, toute interface est mauvaise ou que, en tout cas, aucune n'est optimale. Car à partir du moment où il faut une interface, il y a une rupture du lien direct avec le but qu'on veut atteindre. Pour l'expliquer, je prends l'exemple d'un objet simple : une théière. Si on utilise une théière, ce qu'on veut n'est pas la théière elle-même, mais sentir le thé chaud dans notre corps. Les interfaces se situent ainsi entre nous et notre but : elles sont là pour réaliser ce but. Mais leurs défauts et leur complexité risquent de déprécier l'expérience recherchée.

C'est pour réduire cette dépréciation qu'on travaille sur l'expérience utilisateur et sur l'empathie. On se dote d'outils pour se mettre à la place des utilisatrices ou utilisateurs - avec toute la difficulté que cela représente, parce qu'on ne peut jamais vraiment se mettre à la place de quelqu'un... Disons qu'on essaie ! L'objectif est d'arriver, généralement à travers un nombre incroyable d'échecs, à une forme de simplicité: une interface qui, sans aide extérieure, vous amène au but recherché. Elle le fera en vous proposant des signaux - textes, boutons, couleurs... - que vous êtes susceptible de comprendre ou, en tout cas, qui vous suggèrent des gestes et vous motivent à les accomplir. Les applications qui sortent du lot sont celles qui parviennent à raccourcir le chemin entre vous et le but que vous souhaitez atteindre.

Dans un article récent <sup>1</sup>, vous notez qu'une grande innovation des IA de genre ChatGPT, et une des clés de leur adoption, c'est leur apparente bienveillance. On passe de l'empathie dans le design à une simulation d'empathie par la machine... ChatGPT est paramétré pour faire ressentir une impression de bienveillance, notamment par des artifices de langage et par une mise en scène de l'écriture qui apparaît mot par mot, comme si une personne était là en train de vous écrire. Cette façon de créer la relation, en passant par un simulacre d'humanité qui capte notre

attention et motive notre adoption, constitue un des aspects qui ont permis une acceptation fulgurante de ces programmes.

Cette imitation de bienveillance et d'empathie représente une stratégie commerciale. Il y a une compétition féroce dans ce domaine, et la facilité d'adoption est un enjeu crucial pour les entreprises. On pourrait dire que cette simulation est une technique d'interaction conçue pour tromper l'utilisateur, car feindre l'humanité n'est pas très loyal... Mais, d'autre part, une interface d'usage facile nous procure un certain plaisir, on le voit aussi avec TikTok. Et cela présente plusieurs avantages. Un outil qui serait complètement transparent sur ses limitations, nous répétant sans cesse qu'il n'est qu'une machine et qu'il ne faut rien attendre de lui en termes affectifs, aurait aussi des effets négatifs.

Dans votre article Bienveillant faiseur de mots, vous notez que la simulation d'une relation par ChatGPT provoquerait un rejet immédiat si elle ne sonnait pas si « juste »... Y a-t-il un contre-exemple ?

Un cas très connu est Clippy de Microsoft (1996-2003), un petit personnage en forme de trombone qui surgissait à l'écran et tentait de converser avec l'utilisateur en lui proposant des options. Il n'avait probablement pas été testé suffisamment : c'était tellement

mauvais et il a été tellement rejeté que c'est devenu une source de mêmes (images drôles partagées sur les réseaux sociaux, ndlr) et de moqueries infinies... Est-ce que la simulation d'empathie par les machines représente aujourd'hui une tendance générale ?

Dans une certaine mesure. Le design des sites internet pourrait changer en suivant le modèle ChatGTP : au lieu de

mettre en avant des contenus choisis pour nous intéresser, ils adopteraient l'approche minimaliste de ChatGPT, qui lance un échange en demandant « Comment puis-je vous aider ? » On voit par ailleurs de plus en plus d'appareils émettre des sons qui ne sont pas de simples bruits : ils bipent, parlent, jouent de petites musiques... L'idée, c'est d'intégrer dans les objets qui nous entourent une forme de mini

conversation, des signaux qui poussent à se les approprier plus facilement, à les sentir plus proches de nous. Évidemment, l'empathie reste un trait humain, une machine ne peut pas en avoir. Ce qui est en jeu, c'est l'empathie que les équipes de conception ont essayé de mettre dans ces interfaces. Donc, finalement, il s'agit toujours d'un rapport entre humains.

### Qui a mis des émotions dans la machine ?

Cette figure illustre trois niveaux d'intensité pour les six émotions de base : la joie, la tristesse, la colère, la peur, la surprise et le dégoût. Ce schéma a été utilisé dans le cadre de l'étude Drone in Love : Emotional Perception of Facial Expressions on Flying Robot, menée par une équipe de l'Université Ben Gourion du Néguev en 2021. Elle a exploré comment les expressions faciales affichées sur des drones influencent la perception émotionnelle humaine, suscitant empathie et engagement social. « La folle histoire de Clippy », l'assistant virtuel de Microsoft en forme de trombone, racontée par son créateur Kevan Atteberry. Les étranges promesses des IA sexuelles Une liaison sexuelle avec une créature artificielle peut-elle contenir de l'empathie ? La question traverse le projet de recherche Fucking Tech ! lancé en 2024 pour explorer les enjeux sociaux et culturels de ces partenaires érotiques, ainsi que les possibilités d'un « design des sexualités machiniques » évitant les aspects oppressifs des normes dominantes. Avant de s'attaquer en 2026 aux sexbots (robots sexuels), le projet s'intéresse aux AI Girlfriends, des agentes conversationnelles disponibles en ligne, dotées d'une intelligence et d'une libido artificielles. « Ces applications sont à la croisée de plusieurs champs : sites de rencontre, messageries instantanées, sites porno, avatars de jeux vidéo... Elles incluent des promesses en lien avec l'empathie : les discours marketing les présentent souvent comme des partenaires virtuelles qui prendront soin de vous », constate Anthony Masure, professeur à la Haute école d'art et de design – Genève (HEAD – Genève) – HES-SO et corequérant du projet avec l'enseignant-chercheur Saul Pandelakis. Ces promesses sont-elles tenues ? « Dans les faits, il n'y a pas grand-chose : les dialogues sont stéréotypés. Ces modèles conversationnels ne sont pas faits pour s'intéresser à nous. À la différence de ce qui se passe sur un site de rencontre, où vous commencez par créer votre profil en donnant des informations sur vous (même sommaires), vous y configurez entièrement votre partenaire en décidant de ses caractéristiques physiques et psychologiques. Il y a aussi un biais dans ces applications qui oriente les échanges vers la drague et la sexualité même si vous essayez de parler d'autre chose. » Les questions posées par ces « girlfriends », et les échanges écrits, vocaux et visuels qui s'ensuivent, servent avant tout à alimenter des scénarios sexuels ancrés dans l'hétérosexualité et le regard masculin (les rares AI Boyfriends ne sont pas moins normés). « Il y a une contradiction irrésolue dans ces services entre l'objectification typique des porn tubes et la promesse d'un lien ressemblant à une relation – quelle qu'en soit la forme. » L'informatique affective L'angoisse m'envahit à l'approche d'une échéance. Mon ordinateur le perçoit et décide de m'administrer un remède : un



Hémisphères  
2800 Delémont  
032/ 424 49 00  
<https://revuehemispheres.ch/>

Genre de média: Imprimé  
Type de média: Médias spéciaux  
Tirage: 8'300  
Parution: mensuel

Page: 32,33,34,35  
Surface: 24'511 mm<sup>2</sup>

Ordre: 1073023  
N° de thème: 375009  
Référence:  
57b4d768-26af-4e90-899d-b0b930e6ee35  
Coupure Page: 3/3

flot de lumière bleue et de musique douce. Me revoilà calme, apte à terminer le travail... C'est ainsi que pourrait fonctionner un « environnement de bureau empathique » selon une des expériences menées par Elena Mugellini, professeure responsable de l'institut HumanTech – Technology for Human Wellbeing, avec ses collègues de la Haute école d'ingénierie et d'architecture de Fribourg – HEIA-FR – HES-SO et des confrères des universités de Berne et Fribourg 2. « Cette approche répond à un changement de paradigme. Traditionnellement, on utilisait des appareils informatiques pour accomplir une tâche. Aujourd'hui, nous sommes de plus en plus connectés, et ces systèmes deviennent presque des compagnons technologues. Nous développons une forme d'attachement qui change nos attentes. » Un tel dispositif serait-il compatible avec la loi suisse, qui interdit l'usage de systèmes surveillant le comportement des employées ? « Notre objectif était de tester l'efficacité de l'interaction avec la machine dans une logique d'utilisation personnelle. La question se poserait si une entreprise envisageait de déployer ce genre d'outil auprès de ses collaboratrices et collaborateurs. » Le territoire de l'« informatique affective » a commencé à s'ouvrir dans les années 1990. « Il s'agit de rendre les machines capables de percevoir l'état émotionnel de la personne qui interagit avec elles, de l'interpréter et d'adapter leur réponse », explique Elena Mugellini. Pour déclencher leur « réponse empathique », les machines analysent nos paroles et notre voix, nos expressions faciales, notre regard. Des mesures plus intrusives, telles que la fréquence cardiaque, peuvent compléter l'évaluation. Ces processus pourraient s'implanter dans des voitures semi-autonomes, qui surveilleraient nos émotions pour prévenir des accidents, ou dans des produits tels que Nestore, coach virtuel sur lequel Elena Mugellini a travaillé : « Un projet d'agent conversationnel pour personnes âgées, qui fait des recommandations aidant à garder un style de vie sain en termes de nutrition, d'activité physique ou d'interactions sociales »... Bien sûr, ces dispositifs « ne seront pas un remplacement de l'interaction humaine, mais plutôt un complément ». Avec un atout : à la différence d'un coach humain, son homologue virtuel ne juge pas.

1 Bienveillant faiseur de mots, in Situations 2043, armasuisse Sciences et Technologies, 2024. 2 Dorado A., Daher K., Mugellini E., Lalanne D. & Abou Khaled O., e Effect of Music and Light-Color as a Machine Empathic Response on Stress in Occupational Health, 8th International Conference on Control, Decision and Information Technologies, 2022