



Université catholique de Louvain

Faculté des Sciences Economiques, Sociales et Politiques

Institut Langage & Communication – Pôle Communication

ALGORITHME DE SERVICE PUBLIC ET AUTONOMIE INDIVIDUELLE

Étude des effets de la contrôlabilité d'un système de recommandation sur
l'autonomie de ses usagers

Dissertation doctorale présentée en
vue de l'obtention du grade de
docteur en Information et
Communication par Claes Arnaud

Membres du jury:

Prof. Thibault Philippette (UCLouvain), promoteur et secrétaire du jury

Prof. Pierre Fastrez (UCLouvain), président du jury

Prof. Jean-Samuel Beuscart (Télécom Paris), évaluateur extérieur

Prof. Maude Bonenfant (UQAM), comité d'accompagnement

Prof. Raphaël Jungers (UCLouvain), comité d'accompagnement

Pour m'avoir aidé à réaliser ce qui me semblait impossible, je tiens à remercier...

Mon promoteur, Thibault Philippette, pour sa confiance, son soutien et les nombreux conseils prodigués tout au long de cette thèse.

Mon comité d'accompagnement pour leur patience et leurs conseils avisés.

Mes nombreux professeurs qui ont eu la patience de me guider quand je ne savais pas où aller.

Mes collègues qui ont rempli mes journées d'éclats de rires.

Mes amis pour m'avoir apporté la sécurité nécessaire à mon bien-être et à mon épanouissement.

Ma famille pour avoir supporté mon mauvais caractère.

Et évidemment, Jessi, à qui je dois, avouons-le, toutes mes bonnes idées...

... et sans qui je n'aurais pas eu le courage d'aller aussi loin.

TABLE DES MATIERES

Introduction.....	13
Problématisation.....	13
Plan de la thèse	15
Contexte.....	16
Chapitre 1 — Algorithmes et Pouvoir.....	19
Technologie et pouvoir	20
Le pouvoir comme relation	20
Medium is the message.....	21
Cognition distribuée.....	23
Cognition morale.....	27
Déterminisme et liberté.....	30
Affordance.....	31
Origines du concept.....	31
Définition actualisée	32
De la technologie comme nœud de relations de pouvoir	36
Enjeux des systèmes de recommandation	36
Aux origines de la recommandation automatisée	38
De la surcharge informationnelle à la captation de l'attention.....	45
Pouvoir algorithmique et cyberbalkanisation	50
Du pouvoir algorithmique à sa médiation.....	52
Diversité d'algorithmes.....	52
Diversité des pratiques.....	54
Médiation de l'algorithme.....	56
Chapitre 2 — Interfaces et Autonomie.....	61
Algorithme de service public.....	62
Cadrage général	62
Éditorialisation algorithmique.....	63
De l'éducation à la transparence	65
Médiation de service public	68
Aux origines de l'autonomie	70
Préambule	70

Origines grecques.....	72
L'autonomie morale de Kant.....	73
Romantisme et psychologie humaniste.....	75
Autonomie procédurale.....	79
Authenticité.....	81
Souci de soi.....	84
Pouvoir, domination et liberté.....	85
Culture de soi.....	87
Techniques de soi.....	90
Souci de soi moderne.....	92
Dimensions de l'autonomie.....	97
Autonomie comme compétence créative.....	97
Autonomie comme compétence distribuée.....	100
Autonomie comme attitude critique.....	105
Autonomie et service public.....	105
Designer pour l'autonomie.....	105
Modèle général.....	108
Conclusion intermédiaire.....	111
Chapitre 3 — Méthodologie.....	113
Méthodologie générale.....	114
Préambule méthodologique.....	114
Problématique de recherche.....	117
Structure générale du protocole.....	118
Protocole de conception.....	120
Travaux préexistants sur la contrôlabilité.....	120
Typologie des modalités d'interaction avec le système.....	130
Input.....	131
Processus.....	131
Output.....	132
Explications.....	132
Alveho — Interface utilisateur.....	133
Algorithme de recommandation.....	133
Page d'accueil.....	134
Page de profil.....	137

Fonctionnalités d'administration.....	139
Alveho — Architecture interne.....	139
Récupération des articles.....	140
Gestion des utilisateurs.....	141
Gestion des logs.....	142
Mesure de la durée de l'activité en ligne	142
Démarrage à froid	143
Gestion de l'affichage	144
Protocole expérimental	145
Chapitre 4 — Résultats.....	151
Résultats	152
Évaluation générale de l'activité	152
Compréhension des répercussions	156
Perception des composants	157
Interruption du flux de l'activité	157
Discussion des résultats	159
Retour aux hypothèses.....	159
Freins potentiels à l'autonomie	161
Transparence	161
Gestion de la charge cognitive.....	162
La force de l'habitude.....	162
Leviers réflexifs.....	163
Interruption.....	163
Comparaison	164
Statistiques.....	164
Conclusion.....	165
Limitations.....	165
Perspectives	167
Correction des problèmes d'utilisabilité.....	167
D'une éthique à une esthétique	168
Du souci de soi au souci des autres	170
Leviers d' autolégitimation.....	173
Intrusion et interruption	173
Anthropomorphisation de l'algorithme.....	175

Perspectives en éducation aux médias	177
Cadrage.....	177
Réalisation.....	178
Conclusion	183
Références	191

TABLE DES ILLUSTRATIONS

Figure 1 — Typologie des catégories d’artefacts cognitifs issu de Norman (1992, p.48).....	26
Figure 2 — Représentation graphique du problème de la recommandation collaborative sous la forme d’une matrice utilisateur-objet. Comment prédire la note que Noa va donner au film Encanto en connaissant l’historique de ses jugements ainsi que l’historique des autres membres de la communauté ?	40
Figure 3 — Typologie des techniques de recommandation issue de Burke (2022, p.333)....	42
Figure 4 — Répartition des avantages et inconvénients de chaque technique de recommandation issue de Burke (2022, p.335).....	43
Figure 5 — Capture d’écran de l’interface Netflix en 2008 issu de Lindberg (2019)	46
Figure 6 — Deux exemples d’une page Netflix incluant plusieurs rangées de recommandations issues de Gomez-Urbe et Hunt (2016).....	47
Figure 7 — Modélisation en 3 pôles d’un système de recommandation de service public.....	109
Figure 8 — Structure générale du protocole de recherche.....	120
Figure 9 — Gobo	122
Figure 10 — youchoose.ai.....	123
Figure 11 — Tasteweight issu de Bostandjiev et al. (2012)	124
Figure 12 — Prototype développé par Jin et al. (2018)	125
Figure 13 — SetFusion issu de Parra et Brusilovsky (2015)	126
Figure 14 — Choix d’un algorithme sur MovieLens issu de Ekstrand et al. (2015)	127
Figure 15 — Mock-up développé par Harambam et al. (2019)	128
Figure 16 — Capture d’écran d’un prototype interne à la RTBF - Options de contrôle	129
Figure 17— Capture d’écran d’un prototype interne à la RTBF - Explication des options de contrôle	129
Figure 18 —La page principale d’Alveho	136
Figure 19 — Fenêtre de lecture d’un article sur Alveho	137
Figure 20 —La page de profil d’Alveho	138
Figure 21 — Page d’administration d’Alveho	139
Figure 22 — Guide de la première partie de l’entretien	148

Figure 23 — Évolution du nombre de logs enregistrés pour chaque type de contrôle. « Process controls » réfère aux outils de contrôle de l’algorithme, « Outputs controls » au filtrage des résultats tandis que « Inputs controls » réfère à la gestion de l’historique.....	155
Figure 24 — Capture d’écran du message d’avertissement, issu de Katsaros, Yang et Fratamico (2022)	174

INTRODUCTION

PROBLÉMATISATION

La notion de système de recommandation¹ apparaît dans les années 90 pour désigner un ensemble de technologies qui permettent de suggérer à un utilisateur une sélection d'objets qui sont susceptibles de l'intéresser. Contrairement à des outils de filtrage plus traditionnels qui nécessitent que l'utilisateur formule une requête qui sera utilisée pour inférer ses attentes, ces systèmes de recommandation vont mobiliser des connaissances sur le contenu des objets à recommander, sur l'utilisateur ou sur son environnement social pour déterminer la pertinence des recommandations. A partir de la fin des années 90, ces systèmes sont incorporés dans plusieurs services commerciaux. Depuis 2010, ils sont devenus un élément incontournable du web et de nombreuses plateformes numériques incorporent plusieurs systèmes de recommandation.

Si, au début, ils incarnent une façon confortable et pratique pour filtrer la masse de médias disponibles en ligne, leur usage actuel par des services commerciaux vise de plus en plus à capter l'attention de l'utilisateur. L'intégration à grande échelle de ces systèmes au sein d'infrastructures numériques de plus en plus sophistiquées ainsi que l'intérêt grandissant de médias d'informations pour ces technologies ont suscité de nombreuses craintes relatives à la fragmentation de l'horizon informationnel des utilisateurs de ces services. Parce que cet horizon informationnel est une condition importante à l'exercice de sa citoyenneté (Dufrasne, 2011), ces critiques se sont élargies pour postuler un effet délétère de ces systèmes sur l'espace public et un affaiblissement de l'engagement démocratique.

Ces technologies, initialement pensées pour répondre à des intérêts commerciaux, sont aujourd'hui adoptées par des médias dits de service public. Un important effort de recherche a déjà été déployé pour résoudre la contradiction entre les propriétés de ces dispositifs, calibrés pour capter et conserver l'attention de leurs utilisateurs, et les missions de service public de ces organismes. Ces travaux se concentrent principalement sur la logique de l'algorithme de sorte à garantir une pluralité

¹ Plutôt que de parler d'algorithme de recommandation, nous privilégierons l'emploi de la notion de système de recommandation pour désigner notre objet d'étude. Nous emploierons parfois l'expression d'algorithme de recommandation lorsque nous voudrions mettre l'accent sur l'algorithme plutôt que sur le système dans son ensemble (dont son interface).

de points de vue tout en préservant la capacité de ces dispositifs à distribuer une information jugée pertinente pour l'utilisateur. Si ces recherches contribuent activement à résoudre les enjeux de distributions de l'information, elles laissent de côté les enjeux de développement d'un rapport critique à l'information. L'inappétence pour ces questions s'accompagne d'un désintérêt pour l'interface du système dont l'analyse est souvent réduite à des enjeux de transparence. Or, l'élaboration d'un système de recommandation de service public ne peut se faire sans interroger, de façon précise et méthodique, la façon dont l'algorithme est mis en scène par l'interface du dispositif. C'est cette médiation de l'algorithme qui influera sur la capacité des individus à questionner la validité des décisions prises par ces systèmes.

L'objectif principal de cette thèse visera à explorer, autant que possible, cet angle mort. Plutôt que de nous concentrer sur ce que pourrait être un algorithme « acceptable », « juste » ou « équitable », notre regard portera plutôt sur l'interface de ces dispositifs et son influence sur les conditions d'appropriation de la recommandation par l'utilisateur de ces services. Notre intérêt pour ces questions de médiation contribue à situer nos recherches dans le domaine des sciences de l'information et de la communication. Toutefois, nous avons décidé de formuler notre problématique de recherche, centrée sur la notion d'interface, à la fois comme un *objet d'étude* mais également comme un *projet de design* (Vial, 2021). Dès lors, notre recherche ne sera pas uniquement descriptive mais également prescriptive en ce sens que nous chercherons à élaborer des pratiques de conceptions qui nous apparaîtront compatibles avec les enjeux de service public qui traversent notre recherche. Ce positionnement scientifique, à mi-chemin entre analyse et design, ainsi que notre objet d'étude à la frontière de plusieurs champs disciplinaires, explique l'usage « indiscipliné » que nous ferons de certains cadres théoriques. Nous avons toutefois été attentifs à tempérer nos errances par souci de cohérence globale tout en exposant, de la façon la plus transparente possible, les dimensions normatives et prescriptives de notre approche.

Une intuition très simple est à l'origine de ce projet de design². Lors de notre usage d'objets courants tels qu'une chaise ou un couteau, la configuration de ces objets nous informe sur leur

² Cette intuition est inspirée de propos tenus par Bernard Stiegler lors d'une conférence au séminaire public de la chaire « L'Humain au défi du numérique » du collège des Bernardins en 2016 : « Si je vois une chaise qui me paraît peu solide, je ne m'assois dessus qu'avec prudence. Quelle prudence pour utiliser des algorithmes ? On n'a pas d'intuition directe, la machine nous trompe ». À l'exception de la mention de ces propos par Chandesris (2016), nous n'avons pas trouvé de traces de cette intervention.

fonction et la façon la plus adéquate de les utiliser. Il s'agit de l'idée de base derrière la notion d'affordance. Au-delà de ces considérations fonctionnelles, certaines caractéristiques de ces objets nous informent également sur la confiance que nous pouvons leur accorder ou sur la prudence avec laquelle il est nécessaire de les manipuler. Les craquements inquiétants d'une chaise en bois nous indiquent qu'il vaut mieux ne pas y appliquer tout notre poids. Pour un cuisinier averti, le tranchant de la lame d'un couteau lui indiquera la prudence qu'il devra employer pour effectuer certaines découpes. Un fil émoussé sur lequel il faudra appliquer plus de force se traduira par un risque d'accident plus élevé. Quel pourrait être l'équivalent du craquement d'une chaise pour un algorithme ? Quelles affordances pourraient nous indiquer leur fragilité et nous suggérer une certaine forme de prudence lors de leur utilisation ? Dit autrement, comment encourager l'utilisateur à se préoccuper de son rapport à la recommandation automatisée et à « prendre en main » l'algorithme ?

PLAN DE LA THÈSE

Le premier chapitre de cette thèse vise à clarifier notre cadrage théorique en présentant comment l'algorithme est progressivement devenu une figure de pouvoir au sein de notre écosystème numérique. Pour cela, après avoir explicité ce que nous entendons par « pouvoir », nous mobiliserons deux domaines théoriques qui reposent sur une ontologie relationnelle des rapports entre sujet et objet : d'une part, la théorie de la cognition distribuée à travers les travaux de Hutchins et, d'autre part, les travaux de Verbeek sur la médiation technologique. Ce cadrage nous permettra d'introduire la notion d'affordance qui occupera une place centrale dans notre pratique de conception. La suite du chapitre sera consacrée à la présentation de la notion de systèmes de recommandations, à sa contextualisation historique et à la présentation des principales critiques qui gravitent autour de cet objet. Enfin, nous terminerons ce chapitre en réalisant une revue de la littérature relative à la recommandation automatisée de sorte à positionner notre approche au sein de ce vaste paysage scientifique.

Cette problématique du pouvoir algorithme pose, en creux, la question de l'autonomie de l'utilisateur. Comment définir un usage « libre » ou « indépendant » d'un dispositif qui implique la délégation d'une forme d'autorité ? A l'image du premier chapitre, le deuxième chapitre mettra donc en lien deux notions majeures de notre réflexion : celle d'interface et d'autonomie. C'est par la notion d'algorithme de service public que nous aborderons ces enjeux. Parce que l'emploi de

cette notion d'autonomie peut sembler contradictoire avec notre cadre théorique qui repose sur une conception distribuée de l'agentivité, une partie importante de ce chapitre s'attardera sur les nombreuses significations associées à ce terme. C'est notamment à travers l'exploration des textes de Michel Foucault sur le pouvoir, la liberté et le rapport à soi mais également à travers plusieurs travaux issus de philosophie morale que nous développerons une conceptualisation de l'autonomie qui puisse être opérationnalisée dans le cadre d'une activité de design.

Sur base de ce travail de formalisation, nous pourrons entamer le travail de design prospectif annoncé. Le troisième chapitre est donc méthodologique. Nous reviendrons dans un premier temps sur le protocole général de cette recherche et sur les raisons méthodologiques qui justifient le recours à la conception d'un prototype expérimental. Nous y présenterons ensuite notre question de recherche formulée sur base du travail théorique issu des deux premiers chapitres. La suite de ce chapitre sera consacrée à la présentation d'Alveho, notre prototype expérimental, et à la description de ses fonctionnalités. Enfin, nous conclurons ce chapitre sur la présentation de la méthode employée dans le cadre de l'étude réalisée avec l'aide de notre outil expérimental.

Le dernier chapitre de cette thèse est consacré à la présentation de nos résultats. Ceux-ci nous permettront d'identifier les fonctionnalités qui contribuent à l'autonomie de nos participants mais également de relever plusieurs freins au développement d'un usage critique et réflexif. La suite du chapitre sera consacrée à l'approfondissement de ces résultats et à l'exploration de plusieurs perspectives de recherches qui en découlent. Enfin, nous terminerons notre analyse en présentant les résultats issus d'un atelier en éducation aux médias qui fut réalisé avec Alveho.

CONTEXTE

Cette thèse fut réalisée dans le cadre d'un projet plus large dénommé *Algopinion*. Ce projet, consacré aux enjeux en éducation aux médias de la recommandation automatique d'information, fut financé de 2018 à 2022 par Innoviris, l'organisme public en charge du soutien à la recherche et l'innovation en Région de Bruxelles-Capitale, à la suite d'un appel autour de l'intelligence artificielle, la numérisation et l'automatisation.

Ce projet avait pour préoccupation centrale de dépasser les approches déterministes qui surestiment la force de l'algorithme au détriment de l'agentivité de l'utilisateur (Moeller et Helberger, 2018) et d'évaluer la recommandation automatique comme une composante d'un environnement

médiatique plus large (Bruns, 2019). Ce projet inter-universitaire, réunissant des chercheurs de l'Université Catholique de Louvain (UCLouvain) et de l'Université Saint-Louis Bruxelles (USL-B), s'est également voulu transdisciplinaire en impliquant plusieurs acteurs de la société civile. Pour interroger les enjeux de service public, ce projet pu compter sur le partenariat de la Radio-télévision belge de la Communauté française (RTBF). En ce qui concerne le suivi des enjeux d'éducation aux médias, l'ASBL Média Animation fut le partenaire privilégié du projet. Enfin, le Centre du Traitement Automatique du Langage (CENTAL) de l'UCLouvain fut également impliqué dans certaines parties du projet nécessitant des développements informatiques. Ces trois acteurs furent également impliqués dans la réalisation de cette thèse.

Le projet Algotopinion fut construit autour de trois « work packages » (WP) :

- 1) Observer et analyser les pratiques informationnelles des jeunes sur les réseaux sociaux et collecter leurs opinions sur l'influence de ces systèmes sur leurs comportements ;
- 2) Développer une activité d'éducation aux médias pour encourager un usage critique de ces dispositifs
- 3) Évaluer les modalités d'interfaçage entre l'utilisateur et le système pouvant soutenir l'autonomie de l'utilisateur

Si les résultats du premier WP ont contribué à l'élaboration de notre problématique de recherche, notre intervention se situe plus spécifiquement sur les deux autres volets du projet.

CHAPITRE 1 — ALGORITHMES ET POUVOIR

La problématique de la recommandation automatisée d'information est particulièrement complexe, car elle mêle des questions d'usages avec des considérations techniques pointues. Dès lors, si nous souhaitons nous intéresser aux enjeux de médiation associés à l'usage de ces technologies, il nous faut faire un premier travail de clarification. Ce chapitre a pour objectif général de problématiser notre objet d'étude et de préciser les enjeux de recherche associés.

Il est commun de penser que notre environnement technique est un système neutre et que les résultats de nos interactions avec les dispositifs qui le composent dépendent entièrement de notre volonté. Cette pensée s'exprime souvent à travers l'adage anglais « Guns don't kill people, people kill people ». Plusieurs décennies de recherche en sociologie et en philosophie des techniques nous amènent toutefois à remettre en question ce postulat. Parce qu'il s'agit là d'un fondement de notre réflexion, ce premier chapitre sera consacré, dans un premier temps, à un travail de clarification de nature ontologique. En mobilisant le paradigme de la cognition distribuée ainsi que des travaux issus de la philosophie des techniques, nous poserons le cadre général qui nous permet de comprendre les rapports complexes qui s'établissent dans nos rapports à la technique. Nous nous arrêterons notamment sur les notions d'artefact cognitif et d'interface de Norman (1993) qui nous permettront de détailler la nature du couplage entre un individu et un système technique et ses effets sur les processus cognitifs. Equipé de ce cadrage théorique, nous pourrons plus aisément appréhender les enjeux sociaux relatifs à l'automatisation de la recommandation d'information.

Dans un deuxième temps, à travers un historique succinct de ces technologies, nous réaliserons un travail fondamental de définition de notre objet d'étude. Depuis l'apparition de l'expression « système de recommandation » dans les années 90, cette technologie a connu plusieurs jalons importants dans son histoire qui traduisent plusieurs tournants paradigmatiques notables. Dans le cadre d'un historique de ces étapes cruciales, nous reviendrons notamment sur la plateforme Netflix et l'influence de l'entreprise sur l'interface de ces dispositifs ainsi que sur la place qui est accordée à l'utilisateur dans le processus de recommandation.

Enfin, dans un troisième temps, à la suite de ces deux premières étapes de clarification, nous détaillerons en quoi les systèmes de recommandation peuvent incarner une figure de pouvoir. Si nous présentons les grandes critiques sociales qui gravitent autour de notre objet d'étude, nous

reviendrons également plus en détail sur la littérature récente en sciences sociales afin de construire une représentation plus fine et plus complexe des enjeux posés par ces technologies.

TECHNOLOGIE ET POUVOIR

Le pouvoir comme relation

Parmi l'ensemble des chercheurs et théoriciens que nous citerons tout au long de cette recherche, Michel Foucault est probablement celui dont la pensée a laissé l'empreinte la plus durable tout au long de notre réflexion. Si nous reviendrons plus en détail sur certains de ses travaux dans le deuxième chapitre de cette thèse, une première influence notable concerne la perspective qu'il adopte au sujet du *pouvoir*. Lorsque l'on parle du pouvoir, on pense habituellement au pouvoir de l'État ou d'une institution sur ses constituants, au pouvoir du maître sur l'esclave ou à la domination d'une classe sociale sur une autre. Ces représentations partagent en commun une conceptualisation du pouvoir comme d'une entité qui dispose d'une essence et qu'il est possible de posséder. Foucault s'éloigne de cette perspective pour postuler le pouvoir comme une composante intrinsèque à toute forme de relation : la relation de la mère à l'enfant, de l'enseignant à l'apprenant, de l'employeur à l'employé etc. Il ne s'agit donc pas d'une essence fondamentale mais d'agencements complexes qui articulent pratiques et savoirs et qui traversent l'ensemble des relations sociales. A l'inverse d'une conceptualisation du pouvoir comme d'une structure surplombante qui vient peser sur un groupe d'acteurs, Foucault propose ici une conception immanente à toute dynamique sociale (Foucault, 1999f).

Cette conceptualisation du pouvoir se rapproche plus du domaine du gouvernement que de l'affrontement (Michaud, 2000). L'établissement d'une relation de pouvoir découle de la volonté d'un individu de gouverner un autre selon des objectifs qui lui sont propres en utilisant un certain nombre de tactiques et de stratégies. Gouverner doit se comprendre ici comme une pratique de détermination des conduites. Il ne s'agit pas d'une action immédiate sur l'autre mais d'une action sur l'action. Le pouvoir « opère sur le champ de possibilité où vient s'inscrire le comportement de sujets agissants : il incite, il induit, il détourne, il facilite ou rend plus difficile, il élargit ou il limite, il rend plus ou moins probable ; à la limite, il contraint ou empêche absolument » (Foucault, 1999f, p. 237).

Foucault entend dès lors par gouvernementalité, « l'ensemble des relations de pouvoir et techniques qui permettent à ces relations de s'exercer » (Foucault, 1999f). L'enjeu n'est donc pas de savoir ce

qu'est le pouvoir mais plutôt comment s'exercent ces relations de pouvoir et dans quel domaine. Selon Foucault, un nombre croissant d'objets deviennent l'objet d'une gouvernamentalité dans notre société contemporaine. Lorsque Foucault aborde la question de la folie et de l'enfermement, c'est notamment pour identifier quels systèmes de gouvernement viennent diriger certains domaines de la vie sociale comme la santé mentale et la criminalité.

Appliquer cette conception relationnelle du pouvoir implique de considérer les objets techniques comme des actants à part entière. Cela implique donc de dépasser une approche fonctionnelle de la technologie entendue comme un outil neutre dont les effets dépendent de leur usage.

Medium is the message

Notre objectif n'est pas de retracer l'ensemble de cet héritage scientifique mais plutôt de tracer le cadre général dans lequel viendra s'insérer notre réflexion. C'est donc un peu arbitrairement, probablement à cause de notre ancrage initial dans le champ de la communication, que nous commençons notre itinéraire théorique par les travaux de McLuhan et son fameux aphorisme énoncé en 1964 : « The medium is the message » (McLuhan, 2000). McLuhan emploie ici une notion de média dont il élargit considérablement les limites à tout ce qui prolonge à la fois les capacités cognitives mais également les capacités motrices du corps humain. Typique du style littéraire de McLuhan et de son sens aigu de la formule, l'expression se veut provocante, ce qui ne manqua pas de lui attirer de nombreuses critiques. Toutefois, cette réflexion marque une étape notable dans la critique des technologies et des médias.

A travers cette expression, McLuhan s'oppose à deux erreurs récurrentes qu'il identifie dans la littérature sur les médias et dans le débat public : à la fois l'idée que la technologie est neutre et que le contenu du média doit être la focale principale de l'analyse. Pour ce qui est de la première critique qu'il opère, il l'illustre par les propos suivants tenus par David Sarnoff, fondateur de la NBC :

Nous sommes trop portés à faire de nos instruments technologiques les boucs émissaires des fautes de ceux qui s'en servent. Les réalisations de la science moderne ne sont pas bonnes ou pernicieuses en soi : c'est l'usage que l'on en fait qui en détermine la valeur.
(McLuhan, 2000, p. 29)

A ceci, il répond :

Voilà bien la voix du somnambulisme courant. Imaginons un peu que l'on déclare : « La tarte aux pommes n'est pas en soi bonne ou mauvaise : c'est ce qu'on en fait qui en détermine la valeur. » [...] Ou, plus justement : « Les armes à feu ne sont pas bonnes ou mauvaises en soi : c'est l'usage qui en est fait qui détermine la valeur. » En somme, si les balles atteignent les bonnes victimes, elles sont bonnes. Si le tube-écran projette les munitions qu'il faut aux gens qu'il faut, il est acceptable. Je ne suis pas de mauvaise foi. Il n'y a rien dans les propos de Sarnoff qui résiste à l'analyse, parce que leur auteur ne tient pas compte de la nature des médias, de tous et de chacun des médias. (McLuhan, 2000, p. 29)

McLuhan opère une distinction nette entre d'une part le média comme objet technologique et le contenu qui est l'ensemble des éléments que peut produire l'appareil. Plutôt que de s'intéresser à la production et à la réception de textes, il suggère de réorienter la critique médiatique sur le média en lui-même. Une attention excessive à la production de textes risquerait de masquer les processus par lesquels ces objets affectent la vie individuelle et collective. Une telle ignorance conduit à placer l'individu dans une position de vulnérabilité à l'égard du pouvoir que ces dispositifs peuvent exercer sur lui.

Là où l'approche de McLuhan est particulièrement novatrice, c'est qu'elle dépasse une approche fonctionnelle de l'outil pour s'interroger sur la façon dont l'être humain change en présence de technologies. Le vrai message d'un média ne serait pas le contenu qu'il véhicule mais « les changements d'échelle, de rythme ou de modèles qu'il provoque dans les affaires humaines » (McLuhan, 2000, p. 26). Ceci conduit parfois McLuhan à qualifier les médias d'*environnements*. Cette perspective a notamment inspiré plusieurs écoles de pensées focalisées sur le tissage complexe qui s'établit entre une culture et son environnement technique, notamment l'écologie des médias (Postman, 2000) et la médiologie française (Debray, 1995).

Lors de la publication du livre *The Medium is the Massage : An Inventory of Effects*, une erreur typographique se glisse dans le titre qui aurait dû s'écrire « message » au lieu de « massage ». McLuhan décide de conserver l'erreur qui permet quatre lectures différentes du dernier mot du

titre : Message, Mess Age, Massage, Mass Age³. Chacune de ces lectures symbolise une thématique importante dans les travaux de l’auteur. « Massage » renvoie à l’effet anesthésiant qu’il attribue à certaines technologies médiatiques et à l’industrie du divertissement. « Mess Age » réfère à une forme de chaos social qu’il associe à une évolution technologique irréfrenable. Enfin « Mass Age » renvoie à une autre idée célèbre de McLuhan, celle d’un village planétaire (« global village ») qui émergerait avec le développement des technologies de communication et qui ramènerait l’humanité à une forme de tribalisme. Aujourd’hui encore, chacune de ces préoccupations alimente les débats autour de l’usage de systèmes de recommandation sur lesquels nous reviendrons plus loin dans ce chapitre.

Ce projet d’une vision de la technologie comme extension de nous-mêmes fut notamment explorée dans le domaine des sciences cognitives mais aussi en sociologie et en philosophie des techniques. Les pages qui suivent sont consacrées à la présentation des courants de pensées principaux issus de ces domaines et qui ont influencé notre analyse.

Cognition distribuée

Selon une conception cognitiviste classique, la cognition est avant tout un exercice computationnel dirigé par une unité centrale qui repose sur la manipulation de représentations du monde. Ces représentations sont elles-mêmes acquises par nos sens. Cette perspective, fortement inspirée de la métaphore de l’ordinateur et de la logique formelle, conceptualise donc l’activité mentale comme un calcul de symboles accessibles par la mémoire. A l’opposé de cette posture internaliste, plusieurs travaux suggèrent que la cognition ne se limite pas uniquement à ce qui se passe dans le cerveau :

Where does the mind stop and the rest of the world begin? The question invites two standard replies. Some accept the demarcations of skin and skull, and say that what is outside the body is outside the mind. Others are impressed by arguments suggesting that the meaning of our words « just ain’t in the head », and hold that this externalism about meaning carries over into an externalism about mind. We propose to pursue a third position. We advocate a very different sort of externalism: an active externalism, based

³ Selon le site officiel de la succession de McLuhan (<https://marshallmcluhan.com/common-questions/>)

on the active role of the environment in driving cognitive processes. (Clark et Chalmers, 1998, p. 7)

De ce point de vue, les processus cognitifs ne reposent pas seulement sur la manipulation de représentations internes mais aussi sur les propriétés externes de l'environnement, qu'il soit social, culturel ou matériel. Dès lors, l'unité d'analyse du processus cognitif n'est plus limitée à ce qui se passe au niveau de l'individu et doit prendre en compte le système qu'il compose avec des entités externes qui peuvent être d'autres individus ou des artefacts physiques.

In these cases, the human organism is linked with an external entity in a two-way interaction, creating a coupled system that can be seen as a cognitive system in its own right. All the components in the system play an active causal role, and they jointly govern behavior in the same sort of way that cognition usually does. If we remove the external component the system's behavioral competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head. (Clark et Chalmers, 1998, p. 8)

Une conceptualisation distribuée de l'intelligence prend progressivement de l'ampleur. Minsky (1986), à travers son livre *Society of Mind*, invite à prendre en compte des configurations plus larges dont les agencements permettent l'exercice d'un plus grand niveau d'agentivité et l'émergence de processus cognitifs qui ne seraient pas possibles autrement. « An important property of aggregate systems is that they may give rise to forms of organization that cannot develop in the component parts » (Hutchins, 2001, p. 2069). Hutchins (2001) distingue 3 modalités de distribution de la cognition : une distribution sociale, temporelle mais également une distribution matérielle et environnementale. Outre l'influence notable qu'exercent la culture et l'environnement social sur le développement de la cognition individuelle, Hutchins identifie également que certaines structures sociales permettent l'émergence de propriétés cognitives qui leur sont propres et qui ne peuvent pas être réduites aux propriétés de chaque constituant. La distribution temporelle désigne la dynamique par laquelle des connaissances, des outils, des procédures ou tout autre acquis issus d'un événement cognitif peuvent être transmis au cours du temps au sein d'une même collectivité.

Enfin, la distribution matérielle et environnementale est probablement celle qui nous intéresse le plus ici⁴.

Pour formaliser le rôle joué par l'environnement matériel sur les processus cognitifs, Hutchins reprend le concept d'artefact cognitif développé par Norman (1993). Norman qualifie d'artefact cognitif un « instrument artificiel conçu pour conserver, rendre manifeste de l'information ou opérer sur elle, de façon à servir une fonction représentationnelle » (Norman, 1993, para. 45). Il est courant de penser que l'usage de toute une série d'objets quotidiens amplifie nos capacités cognitives. Toutefois, l'usage d'une calculatrice n'amplifie pas nos capacités arithmétiques. Écrire n'amplifie pas nos compétences mémorielles. Au contraire, le recours à ces objets demande de mobiliser d'autres compétences que celles initialement nécessaires à la réalisation de la tâche. Dès lors, le recours à un artefact cognitif ne conduit pas à une amplification mais plutôt à un processus de réorganisation de compétences cognitives au sein du système composé par l'individu et l'objet (Hutchins, 2001). Hutchins explore notamment l'exemple du nomogramme, d'usage, entre autre, dans le domaine de la navigation. Cet outil graphique, comprenant trois échelles logarithmiques, une pour la vitesse du navire, une pour le temps écoulé et une pour la distance parcourue, permet de connaître facilement l'une des trois valeurs en traçant une droite qui passe par les deux valeurs connues. Le recours à cet outil permet de parvenir à un résultat qui serait également atteignable par une opération purement mentale mais l'exercice serait beaucoup plus difficile. En outre, les compétences mobilisées par l'individu lors de l'usage du nomogramme sont très différentes et capitalisent sur des capacités qui sont plus à l'avantage de l'individu, comme identifier des motifs visuels ou manipuler des objets physiques (Hutchins, 2001). Si la performance globale du système cognitif est augmentée, c'est grâce à cette modification de la nature de la tâche qui est provoquée par le recours à un artefact cognitif. L'artefact agit dès lors comme médiateur à deux niveaux : « dans la perception, entre les transformations du monde et notre détection et interprétation des états créés » (Norman, 1993, para. 34).

Norman (1993) distingue les artefacts capables de représentations superficielles de ceux reposant sur des représentations internes. Si des artefacts comme le boulier ou le livre sont capables de

⁴ Nous reviendrons toutefois sur la distribution sociale dans le chapitre 2 lorsque nous envisagerons la nature distribuée de l'autonomie.

représentations de surface, d'autres ne peuvent avoir que des représentations qui sont stockées à l'intérieur du système. Norman propose la classification suivante en distinguant les artefacts qui maintiennent une représentation et ceux qui permettent leur manipulation :

Form of processing	Representation level	
	Only surface representation	Internal representation
Passive		
Examples	Book. Ruler. Computational tables.	Video and audio tape recorders. Electronic books (dictionaries, translators, Bibles).
Properties	Allows the entry and maintenance of symbols. Cannot modify the symbols according to their content.	Allows the entry and maintenance of symbols. Cannot modify the symbols according to their content.
Requires	A surface representation that can be interpreted by the person and surface operations that can be used to enter information.	An interface to provide a surface representation and surface operations.
Active		
Examples	Slide rule, abacus. Paper & pencil.	Calculator, computer. Clock.
Properties	Allows the entry and maintenance of symbols. Can operate upon and modify the symbols according to their content.	Allows the entry and maintenance of symbols. Can operate upon and modify the symbols according to their content.
Requires	A surface representation that can be interpreted by the person and surface operations that can be used to enter information.	An interface to provide a surface representation and surface operations.

Figure 1 — Typologie des catégories d'artefacts cognitifs issu de Norman (1992, p.48)

Un type d'artefact cognitif particulier, une interface, est nécessaire lorsque l'artefact est caractérisé par des représentations internes que l'individu doit pouvoir traiter. L'interface permet la transformation des représentations internes en représentations de surface que l'individu pourra interpréter et manipuler (Norman, 1993). La conception de cette interface est dès lors centrale dans l'utilisation de ces artefacts :

Pour qu'un artefact soit utilisable, deux conditions doivent être satisfaites : la représentation en surface doit correspondre à quelque chose d'interprétable par l'individu ; les opérations exigées pour modifier l'information dans l'artefact doivent elles aussi pouvoir être exécutées par l'utilisateur. L'interface sert à transformer les propriétés du système représentationnel de l'artefact de façon à établir une correspondance avec les propriétés de l'individu. (Norman, 1993, para. 50)

Norman constate que, tant que le système est de nature mécanique, l'interface est à la fois l'objet sur lequel l'individu opère mais également la représentation de l'état du système. C'est par exemple le cas d'un grille-pain actionné par un levier qui abaisse le pain et allume simultanément la résistance. Il suffit de regarder la position du levier pour en déduire l'état du dispositif. Le développement de l'informatique a contribué à dissocier le lieu de l'action du système représentationnel et, par conséquent, à abstraire l'interface des considérations matérielles (Norman, 1992). Un effort de conception conscient doit donc être déployé pour s'assurer que l'utilisateur du système dispose de tous les signaux nécessaires pour lui communiquer que la tâche demandée a bien été effectuée et garantir la compréhension du dispositif.

Cognition morale

Au-delà des sciences cognitives, une part importante de la critique portant sur les effets de la technologie provient de philosophes comme Jacques Ellul, Gunther Anders, Walter Benjamin, Martin Heidegger ou Karl Jaspers pour ne citer qu'eux. Bien que leurs approches diffèrent sur de nombreux éléments, ils ont toutefois comme point commun de se consacrer aux développements technologiques de l'époque industrielle et des méfaits possibles d'une culture de masse qui commence à émerger. L'individu est généralement entrevu comme un rouage d'un système technique oppressant qui prive le sujet de sa capacité à réaliser sa propre existence. Toutefois, aux alentours des années 70 et 80, sous l'impulsion du domaine des *Science and Technology Studies*, les recherches sur la technique issues de la philosophie et de la sociologie connaissent un important tournant empirique (Achterhuis, 2001). Cette seconde génération s'éloigne de la première en se détachant de la thèse de l'aliénation technologique et d'une approche généraliste trop détachée des réalités techniques, que l'on retrouve notamment dans les travaux d'Heidegger, pour venir approcher la technique dans ses conditions concrètes de production et d'utilisation. Si les considérations issues de la théorie de la cognition distribuée sont plutôt consacrées à la

manipulation de représentations impliquées dans une série de processus cognitifs fondamentaux, la focale d'analyse des travaux issus de la sociologie et de la philosophie des techniques est située à un niveau plus élevé de représentation, celui des valeurs, des normes et, plus largement, de la moralité. Heersmink (2017) parle de cognition morale pour désigner ce niveau d'analyse.

Du côté des STS, les travaux de Latour et la *théorie de l'acteur réseau* en sont probablement les représentants les plus emblématiques. Dans un article intitulé « Where are the Missing Masses? », Latour (1992) nous propose une *petite sociologie des objets de la vie quotidienne*⁵. Il constate que la problématique de la moralité n'est pas seulement une question humaine mais qu'elle transparaît partout dans notre environnement technique. Il prend ainsi l'exemple de la clé berlinoise, une clé caractérisée par la présence de deux pannetons. Chaque panneton est à une extrémité différente de la clé. Lorsque l'on enfonce la clé dans la serrure et que l'on déverrouille la porte avec, un astucieux système d'encoches asymétriques permet de faire en sorte que la clé ne peut être retirée de la serrure. La seule solution, si l'on souhaite récupérer sa clé, consiste d'abord à l'enfoncer complètement dans la serrure pour qu'elle dépasse de l'autre côté de la porte. A ce moment-là, la clé est toujours bloquée dans la serrure mais elle peut maintenant être actionnée depuis l'autre côté. Ce n'est qu'une fois que la porte sera verrouillée de ce côté qu'il sera possible de retirer la clé. De par sa configuration, la clé impose l'impératif « Fermez-la porte derrière-vous » à son propriétaire. De la même façon, le casse-vitesse impose au conducteur de modérer sa vitesse avant de le traverser. Plus tard, Latour prendra également l'exemple des clés d'hôtels auxquelles est attaché un poids suffisamment lourd et encombrant pour décourager le client de l'hôtel de les garder dans sa poche et ainsi garantir leur retour au comptoir de l'accueil (Latour, 2006). Selon cette approche, des artefacts techniques ne sont pas de simples outils mais les médiateurs de notre rapport au réel.

En se basant sur les travaux de Latour mais également d'Ihde (1990), Verbeek développe la notion d'agentivité morale distribuée. Pour ce faire, il s'intéresse à l'imagerie médicale utilisée pendant la grossesse. Verbeek constate que cette technologie reconfigure le rapport moral au fœtus à plusieurs niveaux. Dans un premier temps, l'isolation du fœtus à l'écran, indépendamment du corps de la mère, ainsi que la connaissance de son sexe contribuent à le personnifier bien avant la

⁵ Cette expression sera à la fois le sous-titre de l'article en anglais (Latour, 1992) et, un peu plus tard, le titre du chapitre de son livre *Petites leçons de sociologie des sciences* (Latour, 2006)

naissance. Dans un second temps, en plus de faire du fœtus une personne, l'imagerie médicale le constitue très tôt comme un patient à travers l'identification précoce de maladies. Le recours à l'imagerie médicale fait de l'utérus non pas un lieu intime mais un environnement parfois hostile qu'il faut surveiller. Le contact visuel qui est offert au père facilite son engagement dans le processus. Plutôt que de fournir une fenêtre transparente sur le fœtus, l'imagerie obstétrique transforme activement l'expérience de la grossesse. Ce n'est plus tant « attendre un enfant » que de « choisir un enfant » : choisir de passer le scan, choisir le traitement ou encore, choisir de terminer la grossesse.

Cette approche n'est pas sans rappeler celle de la cognition distribuée où la cognition n'est pas limitée, selon une approche internaliste, à l'activité du cerveau mais plutôt au système composé de l'individu et des artefacts mobilisés pour l'occasion. Ce parallèle est particulièrement visible dans cette citation de Verbeek (2011) lorsqu'il s'intéresse à la figure du *gunman* :

A gun is not a mere instrument, a medium for the free will of human beings; it helps to define situations and agents because it offers specific possibilities for action. A gun constitutes the person holding the gun as a potential killer and his or her adversary as a potential lethal victim. Without denying the importance of human responsibility in any way, we can conclude that when a person is shot, agency should not be located exclusively in either the gun or the person shooting, but in the assembly of both. The English language even has a specific « modern » word for this example: *gunman*, as a hybrid of human and nonhuman elements. The gun and the man form a new entity, and this entity does the shooting. (p.75)

Cette analyse s'applique également à des technologies de plus grande envergure comme le réseau d'approvisionnement électrique qui, en facilitant certains styles de vie nécessitant une importante quantité d'énergie, influence notre façon d'approcher la question environnementale. Notre façon d'agir et de nous rapporter au monde est irrémédiablement médiée par notre écosystème technique. Notre expérience de la conduite d'une voiture est autant liée à l'internalisation d'un code de la route qu'à la structure du réseau routier ou la présence de ralentisseurs contraignant notre style de conduite. Cette analyse conduit Verbeek (2011) à postuler que la capacité d'un individu à développer un jugement moral est profondément connectée à son environnement matériel :

When technologies are used, moral decisions are not made autonomously by human beings, nor are persons forced by technologies to make specific decisions. Rather, moral agency is distributed among humans and nonhumans; moral actions and decisions are the products of human-technology associations. (p.64)

En s'inspirant des travaux de Winner et Latour, Schraube (2009) propose de parler d'action matérialisée pour qualifier la capacité d'un artefact technique à incarner des normes, des valeurs à travers une configuration matérielle qui structure, régule, et discipline la vie humaine, autant au niveau individuel que collectif. Dès lors, la conception d'un objet technique n'est pas une simple production fonctionnelle mais la création de modalités d'existence (Verbeek, 2016).

Déterminisme et liberté

Étant donné ce cadre théorique général que nous avons présenté jusqu'à maintenant, il pourrait nous être reproché d'adopter une posture déterministe à l'égard de la technique. Il s'agit effectivement d'une critique récurrente qui fut notamment adressée à McLuhan. Interrogé sur l'existence d'un sujet souverain, Foucault (1999i) répondra :

Je pense effectivement qu'il n'y pas de sujet souverain, fondateur d'une forme universelle de sujet qu'on pourrait retrouver partout. Je pense au contraire que le sujet se constitue à travers des pratiques d'assujettissement, ou, d'une façon plus autonome, à travers des pratiques de libération, de liberté, comme, dans l'Antiquité, à partir bien entendu, d'un certain nombre de règles, styles, conventions, qu'on retrouve dans le milieu culturel. (p.733)

Nous nous situons ici dans une posture qu'Achterhuis (1998) appelle un *déterminisme minimaliste* qui, contrairement à une approche *maximaliste*, conçoit le système technique comme un facteur transformatif important, mais non déterminant. Si la structure d'un objet contribue à suggérer, avec plus ou moins de force, une utilisation particulière, l'individu reste en mesure de peser dans la relation. Exercer sa liberté ne consiste pas à s'extraire des pressions externes, mais à les mettre en forme de façon consciente. Dès lors, l'idée de liberté ne doit pas être conceptualisée comme une absence d'influences externes mais comme l'espace des possibles composé par notre environnement technique et au sein duquel l'individu s'épanouit (Verbeek, 2011). Dès lors, nous privilégierons les notions de *détermination* (Proulx, 2015) ou d'*inclination* (Ihde, 1990) plutôt que

de déterminisme technologique pour décrire notre approche. Bien que ces deux notions découlent de deux champs disciplinaires différents, la sociologie des usages pour le premier et la philosophie des techniques pour le second, elles décrivent un rapport à la technique de l'ordre de la conduite plutôt que du contrôle.

L'ensemble du deuxième chapitre sera dédié à l'approfondissement d'une telle conception de la liberté et de l'autonomie. Nous reviendrons donc en profondeur sur cette question du sujet souverain dans le deuxième chapitre de cette thèse, mais nous souhaitons d'abord insister sur la place fondamentale de la liberté individuelle dans notre approche, tout en refusant également cette idée d'un sujet souverain. En regard de cette subtilité, le concept qui nous semble le plus opérationnel pour nous permettre d'observer comment une technologie, en tant qu'action matérialisée, influence des conduites est probablement celui d'*affordance*.

AFFORDANCE

Origines du concept

L'approche de Gibson s'oppose à l'approche dominante de la perception dans les années 60 qui l'identifie comme un processus d'inférence de représentations médiatisées sur base de connaissances préexistantes. Selon Gibson, l'action peut reposer sur une perception directe des choses, sans processus d'inférence préalable (Davis, 2020). La notion d'affordance est alors centrale dans sa réflexion. La définition canonique qu'il en donne dans son livre *The Ecological Approach to Visual Perception* est la suivante :

The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, but the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment. (Gibson, 2015, p. 119)

Depuis cette conceptualisation initiale, la notion d'affordance fut mobilisée dans de nombreuses disciplines différentes. L'interprétation la plus influente qui en fut faite et qui irrigue encore aujourd'hui une grande majorité des recherches sur le sujet est probablement celle proposée par Norman dont nous avons déjà cité les travaux précédemment :

The term affordance refers to the relationship between a physical object and a person (or for that matter, any interacting agent, whether animal or human, or even machines and robots). An affordance is a relationship between the properties of an object and the capabilities of the agent that determine just how the object could possibly be used. A chair affords (« is for ») support and, therefore, affords sitting. Most chairs can also be carried by a single person (they afford lifting), but some can only be lifted by a strong person or by a team of people. If young or relatively weak people cannot lift a chair, then for these people, the chair does not have that affordance, it does not afford lifting. (Norman, 2013, p. 11) ⁶

L'affordance n'est pas tant la propriété technique d'un objet que le produit d'une relation entre un artefact et un individu. Cette dimension est déjà bien présente chez Gibson qui définit l'affordance comme des opportunités d'actions découlant des relations entre des propriétés de l'objet et des propriétés du sujet (Gibson, 2015). En utilisant la notion d'opportunité, il met en évidence la nature potentielle de ces actions dont l'actualisation dépend à la fois des propriétés de l'objet et des compétences de l'individu. Norman soulignera une nouvelle fois l'importance des compétences individuelles lorsqu'il aborde la notion d'artefact cognitif : « La perception de l'artefact et son usage varient en effet en fonction de la nature de la tâche et du niveau d'expertise et d'habileté de la personne » (Norman, 1993, para. 22).

Cette notion d'affordance est un des piliers de ce qu'il appelle un *human-centered design*. Le travail de conception doit se mettre au service des besoins de l'utilisateur tout en étant capable de s'ajuster aux capacités des utilisateurs. Le rôle du designer consiste donc à guider l'utilisateur à travers la mise en place des opportunités d'actions adéquates. A ceci s'ajoute, dans le cas d'outils informatiques, le besoin d'établir une correspondance efficace entre le système représentationnel en surface et les transformations des représentations internes du dispositif.

Définition actualisée

Depuis ses débuts, la notion d'affordance fut mobilisée dans de nombreux contextes différents et notamment, depuis l'usage qu'en fait Norman, pour l'analyse d'environnements numériques (voir

⁶ Norman publie *The Psychology of Everyday Things* in 1988. Nous mobilisons ici la version révisée et étendue de ce livre, publiée en 2013 et renommée *The Design of Everyday Things*.

par exemple : boyd, 2010 ; Bucher et Helmond, 2018 ; Gillespie, 2010). La démultiplication des usages et l'étalement du concept ont naturellement contribué à l'apparition de nombreuses critiques. Outre la volatilité du concept qui conduit à des interprétations parfois contradictoires, Davis (2020) identifie deux critiques majeures qui sont formulées à l'égard du concept d'affordance : la binarité du concept ainsi que sa rigidité.

La première critique réfère à l'usage binaire du concept dans de nombreuses recherches qui cherchent à identifier comment certaines propriétés « affordent⁷ » ou « n'affordent » pas. Selon Davis (2020), il est crucial de développer un vocabulaire qui permette de saisir toute la nuance avec laquelle certaines configurations techniques peuvent suggérer des comportements et qui permette de dépasser une opposition simpliste entre autorisation et interdiction. De plus, trop souvent, la notion d'affordance est envisagée comme statique et monolithique ce que Davis (2020) trouve assez paradoxal sachant que la conceptualisation initiale du terme vise à garder une certaine fluidité à l'égard de la relation entre le sujet et l'objet. Alors que la plupart des analyses repose sur l'identification de propriétés figées, le concept doit pouvoir capturer la variabilité inhérente aux multiples contextes sociaux et à la diversité des profils d'utilisateurs.

Pour répondre à ces critiques, Davis (2020) propose, dans son livre sous-titré *The Power and Politics of Everyday Things*, un modèle qui nous invite à passer d'une perspective du « quoi » (qu'est-ce qui est suggéré comme comportement ?) à une perspective du « comment » (comment est-ce qu'un comportement est suggéré ?). Pour échapper à la critique de la binarité, il propose un modèle qui demande de distinguer les affordances qui *demandent* (request), *exigent* (demand), *encouragent* (encourage), *découragent* (discourage), *refusent* (refuse) ou *autorisent* (allow).

Ce modèle, qui divise la notion d'affordance en différents niveaux de pressions sur l'utilisateur, n'est pas sans rappeler la façon dont Foucault conceptualise l'exercice du pouvoir comme une conduite sur les conduites. Pour la forme et la clarté de notre argument, nous replaçons une nouvelle fois sa citation sur le pouvoir :

⁷ Contrairement à l'anglais, le français n'a pas un verbe spécifique qui conserve son affiliation directe à la notion d'affordance. Bien que cela soit assez peu élégant, nous utiliserons parfois l'anglicisme « afforder » par facilité.

Il opère sur le champ de possibilité où vient s'inscrire le comportement de sujets agissants : il incite, il induit, il détourne, il facilite ou rend plus difficile, il élargit ou il limite, il rend plus ou moins probable ; à la limite, il contraint ou empêche absolument. (Foucault, 1999f, p. 237)

Une demande suggère une modalité d'action sans pour autant l'imposer. Le casse-vitesse que nous avons déjà analysé avec Latour, demande à son utilisateur de s'arrêter. Refuser cette demande reste possible même si cela génère un risque important pour le véhicule et le conducteur. Une exigence conditionne l'usage qu'il sera possible de faire de l'artefact. Les plateformes de publication d'articles scientifiques exigent de l'utilisateur qu'il paie son accès ou qu'il passe par une affiliation institutionnelle pour accéder au contenu. Même dans le cas d'une exigence, il est toujours possible pour l'utilisateur de contourner ce qui lui est imposé. Toutefois, cela demandera un effort plus important et une certaine imagination.

Si une demande ou une commande provient immédiatement du dispositif, un encouragement, un découragement ou un refus désigne une réaction du dispositif à l'égard d'un comportement de l'utilisateur. Une assiette plus petite encourage l'utilisateur à manger de plus petites quantités. L'infrastructure de Facebook encourage la création de contenus et la connexion avec d'autres utilisateurs à travers la mise en évidence de certaines fonctionnalités (like, partage, tagging, etc.). Dans ces cas-là, les comportements encouragés sont plus aisés à réaliser et l'interaction avec l'artefact est fluide. A l'inverse, la mise en place d'obstacles permet de décourager certains comportements. Ces obstacles ne sont pas nécessairement une fonctionnalité concrète mais peuvent découler de normes sociales. Davis (2020) observe ainsi que les normes qui gravitent autour de l'usage d'Instagram découragent la production de plus d'un contenu par jour pour ne pas obstruer le fil d'actualité des abonnés. Aucune limitation technique n'est mise en place pour imposer ce comportement qui découle avant tout d'un code de « bonne conduite » informel.

Le design d'un dispositif peut également viser à interdire complètement certains comportements. L'interface utilisateur de twitter interdit la rédaction de messages trop longs. De la même façon qu'avec les exigences, il est tout à fait possible de contourner ces interdictions. Pour contourner la limite de signes imposée par twitter, il est fréquent de réaliser un « fil » de tweets qui comprend plusieurs tweets consécutifs dont l'ordre de référencement détermine l'ordre de lecture. Enfin, l'autorisation occupe un statut un peu particulier dans le modèle de Davis (2020) :

Allow applies to bids placed by technologies and to bids placed on technologies. Allow is distinct from other mechanisms of affordance due to its neutral intensity and multidirectional application. A user may take a line of action, but there is no pressure to do so, and there are no significant obstacles in the way. Allow is like a fork in the road. A traveler may just as easily opt for one route as another. The traveler is not faced with enticements from any direction, and the traveler does not need to overcome any extra blockades to access the pathways. (p.38)

Plutôt qu'un dispositif à la structure rigide, ce modèle est avant tout un outil d'analyse qui se veut flexible. Un casse-vitesse ou un *paywall* peut être identifié comme une demande par l'un ou comme une exigence par l'autre.

These mechanisms are neither rigid nor exhaustive, but rather serve as analytic pegs that transpose structure onto subject-artifact relationships, which are at once nebulous and identifiably patterned. Indeed, features can rest ambiguously between categories and slip from one category to another. The mechanisms are thus reference points along a gradated continuum. (Davis et Chouinard, 2016, p. 242)

Pour garantir la flexibilité du modèle, Davis (2020) complète ces mécanismes par une série de 3 conditions : perception, dextérité et signification culturelle et institutionnelle. Là où les mécanismes permettent de déterminer comment est suggéré un comportement, les conditions précisent pour qui et dans quelles circonstances ces affordances sont plus à même de s'actualiser. L'évaluation d'une affordance implique donc de déterminer si l'individu perçoit les fonctionnalités du dispositif, s'il dispose de la dextérité nécessaire (qu'elle soit cognitive ou physique) pour le manipuler et si l'usage qui est fait de l'artefact est légitimé par son environnement sociale et culturel (Davis et Chouinard, 2016).

Pour les mêmes configurations matérielles, plusieurs individus peuvent percevoir des affordances différentes. Pour un expert en informatique, un système particulier peut être parsemé d'opportunités et de faiblesses qu'il pourrait exploiter. A l'inverse, un grand nombre d'utilisateurs de plateformes commerciales sont souvent limités par les paramètres par défaut et ne perçoivent pas les éléments sur lesquels ils peuvent agir. Prendre en compte les enjeux de dextérité implique de s'interroger sur certaines pratiques de conception qui se focalisent sur l'individu « moyen » et qui occultent les difficultés rencontrées par les utilisateurs qui ne correspondent pas à ce standard.

Cette divergence peut tout autant être une question de handicap physique, sensoriel ou cognitif mais aussi, de façon beaucoup plus commune, une question de compétence et de maîtrise. Enfin, la position de l'individu au sein d'une structure sociale influence également son rapport à différentes configurations matérielles. Un cordon policier peut incarner une exigence plus ou moins forte en fonction de son rapport à l'autorité, de son statut professionnel et de la connaissance de tout un registre culturel.

De la technologie comme nœud de relations de pouvoir

En conceptualisant la technique comme une action matérialisée, nous nous éloignons quelque peu de théories comme celles de l'acteur-réseau en ne considérant pas la technologie comme un *actant* à part entière. Dès lors, si nous inscrivons les artefacts techniques que nous étudions ici dans une relation de pouvoir, ce n'est pas tant comme participant actif mais plutôt comme intermédiaire ou comme médiateur d'une relation entre l'utilisateur et le concepteur du dispositif (qu'il soit un individu, une institution, une collectivité etc.).

Si un artefact technique exerce du pouvoir sur son utilisateur, c'est à travers le champ lexical de la conduite ou du gouvernement que nous avons préféré le qualifier. La notion d'affordance, surtout selon le modèle proposé par Davis (2020), s'articule particulièrement bien avec cette problématisation. Cette conduite des comportements ne concerne pas uniquement l'exercice de compétences cognitives fondamentales mais également le développement de représentations, d'attitudes et d'un positionnement moral. Le reste de ce chapitre sera dédié à la transposition de ce cadrage général à la problématique des algorithmes et, plus précisément, des systèmes de recommandation.

ENJEUX DES SYSTÈMES DE RECOMMANDATION

Les technologies qui nous intéressent plus particulièrement dans le cadre de cette thèse ont pour fonction principale d'ordonnancer automatiquement l'information. Rieder (2020) emploie l'expression de *machines d'ordre* (*engines of order*) pour qualifier ce vaste écosystème de techniques qui sélectionne, indexe, classifie et recommande de l'information. L'adoption croissante de ces technologies au sein de l'écosystème numérique a contribué à faire de ces dispositifs algorithmiques une nouvelle figure de pouvoir qui s'exprime principalement à travers leurs opérations de classification et leur gestion de la visibilité en ligne (Bucher, 2018).

En devenant une figure de pouvoir, ces technologies sont également devenues l'objet de nombreuses critiques. La littérature qui s'est composée autour de ces objets techniques s'est principalement consacrée aux conséquences sociales de l'adoption à grande échelle de systèmes algorithmiques. L'analyse prend souvent une focale macroscopique en se positionnant à l'échelle de grandes structures commerciales ou institutionnelles. En s'insérant généralement dans la continuité de la critique d'une société du calcul (Cardon, 2015), ce pouvoir algorithmique est étudié sous l'angle des normes et des valeurs qu'il véhicule et renforce (Bucher, 2018). Cette dynamique du pouvoir algorithmique peut être résumée par ce que Rouvroy et Berns (2013) qualifient, en s'inspirant des travaux de Foucault, de *gouvernementalité algorithmique* :

Par gouvernementalité algorithmique, nous désignons dès lors globalement un certain type de rationalité (a)normative ou (a)politique reposant sur la récolte, l'agrégation et l'analyse automatisée de données en quantité massive de manière à modéliser, anticiper et affecter par avance les comportements possibles. (p.173)

Une fois réduit à ses traces numériques, l'homme devient un document comme les autres (Ertzscheid, 2009). En prenant les relations sociales comme objet de classement et d'organisation, ces machines d'ordre informationnel deviennent des machines d'ordre social (Rieder, 2020). Dit autrement, l'informatique se fait *sociomatique* (Fastrez et Philippette, 2017).

Plutôt qu'un objet figé, c'est un vaste répertoire de techniques qui se cache derrière la notion d'algorithme employée dans la littérature critique en sciences sociales. Parce que nous ne serons pas en mesure d'en faire une analyse systématique, les quelques pages qui suivent visent à restreindre notre objet d'étude à ce qui est communément réuni sous l'appellation de systèmes de recommandation.⁸ Une fois que nous aurons opéré cette délimitation technique, nous glisserons vers la littérature critique qui s'est développée autour de ces dispositifs pour appréhender les formes de pouvoir qui sont à l'œuvre et les enjeux concomitants d'accès à l'information.

⁸ Nous renvoyons vers les travaux Rieder (2020) pour ceux qui souhaiteraient un aperçu de l'entreprise qui se cache derrière une analyse approfondie de cette écologie technique.

Aux origines de la recommandation automatisée

Bien que la notion de système de recommandation n'apparaisse que dans les années 90, on peut déjà identifier une série de travaux préoccupés par l'automatisation de l'accès à une information pertinente. L'essai de Vannevar Bush « As We May Think », publié en 1945, marque les débuts du domaine de la Recherche d'Information (RI) et on parle alors principalement de filtrage plutôt que de recommandation (Bush, 1945). En 1975, Salton *et al.* (1975) développent un modèle d'indexation automatique de texte qui repose sur la transformation du document en un vecteur dont chaque élément, correspondant à une caractéristique du document, est pondéré selon son importance. La technique la plus commune consiste, pour chaque terme du document, à calculer le produit de la fréquence du terme dans le document par la fréquence inverse du document qui indique la fréquence du terme dans le corpus⁹. La valeur obtenue pour chaque terme compose le vecteur du document. Sur base du vecteur de deux documents différents, il est ensuite possible de calculer le cosinus de leur angle pour obtenir un indice de similarité. Nous en sommes alors au tout début de ce qui s'appellera plus tard le *text-mining* ou la fouille de texte.

Ces approches sont progressivement regroupées sous l'appellation de *content-based filtering* ou filtrage basé sur le contenu pour désigner le fait que ces techniques se basent sur la nature et la structure des objets à filtrer pour les organiser les uns par rapport aux autres et ainsi définir des mesures de proximité relatives au sein d'un même catalogue de textes. Il faudra attendre plusieurs dizaines d'années et l'amélioration des techniques de traitement automatique du langage pour permettre à ces techniques d'atteindre leur maturité. Ceci s'explique notamment par la complexité qu'implique de comprendre la structure d'un domaine ainsi que d'identifier la nature des relations entre les caractéristiques d'un objet et les préférences d'un utilisateur. Le *Music Genome Project*, conçu par Glaser *et al.* (2006) à la fin des années 90, est un exemple abouti de ce type d'approche. Chaque son est décomposé en un vecteur dont chaque valeur correspond à un « gène ». Chaque gène est une caractéristique d'une musique (l'orchestration, le genre, la durée etc.) et chaque vecteur comprend entre 150 et 450 gènes. Un algorithme d'appariement permet ensuite d'associer un utilisateur à d'autres musiques en fonction des gènes des musiques qu'il a déjà appréciées.

⁹ Pour un terme t d'un corpus c , la fréquence inverse du document s'obtient en calculant le logarithme du nombre total de documents de c divisé par le nombre de documents contenant t .

Jusqu'à la fin des années 70, ce filtrage dépend en grande partie de la nature de la requête et des connaissances disponibles sur le domaine d'application. On ne parle pas encore de personnalisation. En même temps que les recherches sur le filtrage basé sur le contenu se poursuivent, Rich (1979) développe *Grundy*, un libraire informatique capable d'interroger un individu pour inférer ses préférences et lui recommander des livres. Sur base des réponses aux questions, Grundy attribue à l'utilisateur une série de stéréotypes auxquels sont reliés des thématiques et des genres littéraires qui permettront d'identifier les livres à recommander. Bien que la méthode soit quelque peu primitive et que tous les utilisateurs associés à un même stéréotype reçoivent les mêmes recommandations, Grundy marque un changement de paradigme important. Il s'agit d'une des premières tentatives de personnalisation du filtrage en fonction du profil de l'utilisateur tout en mobilisant une base de connaissances sur laquelle faire reposer les inférences du système.

En se basant sur l'état de l'art de l'époque en psychologie, Rich est toutefois le premier à reconnaître que l'utilisateur n'est pas toujours la source d'information la plus fiable pour éclairer ses propres intentions et que cette approche est limitée par la portée de ses capacités d'introspection (Rich, 1979). On voit déjà apparaître le dilemme qui anime encore la communauté scientifique entre le recours à des signaux implicites ou explicites pour identifier précisément les préférences de l'utilisateur. Si des dispositifs comme Grundy reposent sur des signaux explicites récoltés à travers un mécanisme de questions/réponses, les partisans d'une approche behavioriste prescrivent plutôt d'observer directement le comportement de l'utilisateur.

En 1992, Goldberg *et al.* (1992) développent *Tapestry*, un outil de tri de documents textuels, qu'ils décrivent alors comme un outil de *filtrage collaboratif*. En plus de permettre de composer des requêtes portant sur le contenu du texte, Tapestry permet également de formuler des requêtes portant sur des annotations laissées par les autres utilisateurs. L'annotation la plus banale est une variable booléenne indiquant si l'on a apprécié ou pas le document. Le système peut aussi prendre en compte des signaux d'appréciation implicites comme le nombre de réponses d'un mail. Si nous prenons pour exemple une recherche dans une base de données d'articles académiques, une requête-type peut ainsi inclure les mots clés « algorithmes » et « recommandation » mais aussi préciser un nombre minimal de citations pour réduire le nombre de résultats. A ce moment-là, le

filtrage collaboratif n'est pas automatisé et demande à l'utilisateur de formuler des requêtes plus ou moins complexes.

La notion de filtrage collaboratif gagne progressivement en popularité. En 1994, Resnick *et al.* (1994) publient le premier système de filtrage collaboratif automatisé permettant de trier des articles issus du réseau *Usenet*. Par après, d'autres systèmes collaboratifs apparaissent pour filtrer des artistes musicaux (Shardanand et Maes, 1995) ou des films (Hill *et al.*, 1995). Le système proposé par Resnick *et al.* (1994) repose également sur la production d'annotations de la part des utilisateurs du service. Ici, l'appréciation d'un document s'exprime par un score sur une échelle de 1 à 5. L'ensemble de ces annotations peuvent être représentées sous la forme d'une matrice utilisateur-objet. Une cellule vide indique qu'un utilisateur n'a pas encore attribué de score à l'élément correspondant. Le tableau ci-dessous représente une version simplifiée d'une telle matrice pour l'annotation de films.

	Le Seigneur	Green Book	Les Heures	Kaamelott	Encanto
Basile	1	5	5	1	4
Louise	3	5	4	1	1
Emile	5	1	2	4	4
Noa	3	4	4	1	?

Figure 2 — Représentation graphique du problème de la recommandation collaborative sous la forme d'une matrice utilisateur-objet. Comment prédire la note que Noa va donner au film Encanto en connaissant l'historique de ses jugements ainsi que l'historique des autres membres de la communauté ?

L'approche proposée par Resnick *et al.* (1994) marque un tournant intéressant dans le domaine. La difficulté technique ne repose pas ici sur une mesure de similarité entre deux contenus ou sur l'élaboration d'une base de connaissances complexe mais sur un travail de prédiction. En connaissant l'historique des préférences d'une communauté d'utilisateurs, comment prédire précisément le score que va attribuer un utilisateur à un objet qu'il n'a pas encore évalué afin de ne lui proposer que des contenus dont le score prédit est élevé ? Pour résoudre ce problème, Resnick *et al.* (1994) proposent alors de se baser sur l'heuristique suivante : des individus qui se sont mis d'accord par le passé sont susceptibles de l'être à nouveau à l'avenir. Pour implémenter cette

heuristique, le système calcule d'abord un coefficient de corrélation qui permet de déterminer à quel point deux utilisateurs tendent à se mettre d'accord sur leurs notations. Pour prédire la note qu'un individu attribuera à un élément du catalogue, il faut alors calculer la moyenne de l'ensemble des notes attribuées par ses pairs à l'élément en question tout en pondérant chaque notation par le coefficient de corrélation correspondant. Si nous reprenons l'exemple ci-dessus, le profil de Louise est proche de celui de Noa ce qui signifie que l'opinion de Louise sur le film *Encanto* aura plus de poids dans le score final. Il est donc fort probable que la note qu'attribuera Noa sera proche de la note déjà attribuée par Louise. A l'inverse, le profil d'Emile est assez éloigné de celui de Noa. Dès lors, son avis aura moins d'importance sur la prédiction finale. En conservant cette heuristique générale, de nombreuses recherches se sont par la suite penchées sur le calcul de ce score de corrélation ainsi que sur des méthodes alternatives d'agrégation de ces scores.

En 1997, Resnick et Varian (1997) proposent de restreindre l'usage de la notion de filtrage collaboratif et d'adopter le terme plus général de *recommender system* (système de recommandation) afin d'inclure des systèmes qui ne sont pas explicitement collaboratifs. L'usage de ce terme permet également de sortir du registre lexical du filtrage et ainsi d'ouvrir le domaine à des techniques qui se focalisent sur la sélection de contenus pertinents en plus de l'exclusion d'éléments inadéquats. En 2002, Burke (2002) distingue encore les systèmes de recommandation des systèmes de recherche d'information qui, contrairement au premier, sont censés renvoyer l'ensemble des résultats correspondant à la requête formulée par l'utilisateur. Toutefois, il constate, déjà à l'époque, que cette frontière est amenée à s'effacer avec l'intégration progressive de critères de pertinence permettant d'ordonner les résultats de façon plus précise et ainsi améliorer l'utilité des résultats. A cette époque, il pense notamment au critère d'autorité de l'algorithme PageRank de Google (Brin et Page, 1998).

Toujours dans le même article, Burke (2002) identifie 5 logiques de recommandations principales. Les deux premières concernent la recommandation collaborative et la recommandation basée sur le contenu dont nous venons de retracer l'historique. A ces deux catégories, il ajoute:

- Recommandations démographiques : comme son nom l'indique, ce type d'algorithme va se baser sur des informations démographiques détenues par le système au sujet de l'utilisateur pour déterminer les contenus les plus pertinents à lui recommander.

- Recommandations basées sur une fonction d'utilité : cette technique repose sur l'utilisation d'une fonction d'utilité qui permet de calculer la pertinence d'un objet en fonction de ses caractéristiques. Cette fonction d'utilité étant dépendante de l'utilisateur, c'est souvent lui qui doit la spécifier pour le système.
- Recommandations basées sur un modèle de connaissances : les techniques associées à cette dernière famille reposent sur une base de connaissances qui permet de définir comment certains objets sont plus à même de répondre à des besoins spécifiques. En mobilisant cette base de connaissances, cette technique permet de déterminer par inférences logiques quelles sont les objets qui correspondent le mieux aux besoins de l'utilisateur.

Technique	Background	Input	Process
Collaborative	Ratings from U of items in I .	Ratings from u of items in I .	Identify users in U similar to u , and extrapolate from their ratings of i .
Content-based	Features of items in I	u 's ratings of items in I	Generate a classifier that fits u 's rating behavior and use it on i .
Demographic	Demographic information about U and their ratings of items in I .	Demographic information about u .	Identify users that are demographically similar to u , and extrapolate from their ratings of i .
Utility-based	Features of items in I .	A utility function over items in I that describes u 's preferences.	Apply the function to the items and determine i 's rank.
Knowledge-based	Features of items in I . Knowledge of how these items meet a user's needs.	A description of u 's needs or interests.	Infer a match between i and u 's need.

Figure 3 — Typologie des techniques de recommandation issue de Burke (2022, p.333)

Ces 3 logiques de recommandation (démographiques, basées sur une fonction d'utilité ou basées sur un modèle de connaissance) reposent toutes sur la production, soit de la part de l'utilisateur ou du développeur de l'algorithme, d'un certain savoir sur les relations entre des besoins, les caractéristiques des objets à recommander et les caractéristiques des utilisateurs. Etant donné que ce travail est bien souvent manuel, les systèmes collaboratifs ou basés sur le contenu sont généralement les plus répandus car ils ne nécessitent pas cet investissement initial. Toutefois, ils

demandent de grandes quantités de données pour fonctionner. Le tableau ci-dessous, élaboré par Burke (2002), nous donne une vue d'ensemble sur les avantages et inconvénients de chaque technique.

Technique	Pluses	Minuses
Collaborative filtering (CF)	A. Can identify cross-genre niches. B. Domain knowledge not needed. C. Adaptive: quality improves over time. D. Implicit feedback sufficient	I. New user ramp-up problem J. New item ramp-up problem K. 'Gray sheep' problem L. Quality dependent on large historical data set. M. Stability vs. plasticity problem
Content-based (CN)	B, C, D	I, L, M
Demographic (DM)	A, B, C	I, K, L, M N. Must gather demographic information
Utility-based (UT)	E. No ramp-up required F. Sensitive to changes of preference G. Can include non-product features	O. User must input utility function P. Suggestion ability static (does not learn)
Knowledge-based (KB)	E, F, G H. Can map from user needs to products	P Q. Knowledge engineering required.

Figure 4 — Répartition des avantages et inconvénients de chaque technique de recommandation issue de Burke (2002, p.335)

En dehors du monde académique, la majorité des systèmes de recommandation sont des hybrides qui capitalisent sur un mélange de plusieurs logiques de recommandations pour compenser leurs faiblesses respectives. Plusieurs solutions existent pour composer ces hybrides. La technique la plus simple consiste à agréger les scores attribués par chaque sous-système dans une combinaison linéaire (Burke, 2002). L'ajout de poids à chaque score permet d'équilibrer l'importance relative de chaque technique lors de la composition du score agrégé¹⁰. Parce que l'exploration des solutions d'hybridations de ces algorithmes dépasse le cadre de cette thèse, nous renvoyons le lecteur aux

¹⁰ Nous opterons pour une solution similaire lorsque nous développerons notre propre outil expérimental dans le chapitre 3.

travaux de Burke (2002). Alors que des systèmes issus de la recherche sont rarement utilisables en conditions réelles, ces solutions hybrides sont souvent plus adaptées aux problématiques rencontrées par des services commerciaux. Les évolutions récentes dans le domaine du traitement de l'information et notamment l'apprentissage profond ont contribué à élargir le registre et la complexité des informations pouvant alimenter ces techniques de recommandation. Les méthodes d'hybridations se sont aussi complexifiées avec le temps. Toutefois, ces quelques grands principes restent toujours d'actualité et continuent de structurer le domaine.

Les premières solutions commerciales apparaissent dès 1996 (Konstan et Riedl, 2012). A la fin des années 90, le système de recommandation devient une composante essentielle des plateformes de commerce en ligne comme Amazon (à l'époque, vente de livres en ligne) ou CDNow.com (vente de CDs) mais aussi sur des sites comme MovieFinder.com (base de données de films) (Herlocker *et al.*, 1999). Cette importance commerciale a contribué à l'émergence d'un domaine de recherches entièrement consacré à ces technologies qui compte aujourd'hui sa propre conférence internationale (« ACM Conference Series on Recommender Systems¹¹ ») ainsi qu'un nombre important de productions scientifiques et de cours universitaires (Konstan et Riedl, 2012). Jusque-là, la majorité de la recherche sur le sujet se consacrait principalement aux techniques collaboratives. Si la mesure d'appréciation repose souvent sur une cotation explicite, des mesures de satisfaction implicites commencent à être envisagées comme le temps passé à consulter un objet (Herlocker *et al.*, 1999).

Étant donné que le problème de la recommandation est formalisé comme un problème de prédiction, l'évaluation des résultats repose principalement sur la mesure de leur précision à travers des méthodologies d'évaluations « offline » (Konstan et Riedl, 2012). Ces évaluations reposent sur l'utilisation d'une base de données dans laquelle sont enregistrés des logs d'interactions réelles avec des objets à recommander. On peut notamment y retrouver des scores d'appréciation attribués par des utilisateurs à l'image de la matrice présentée ci-dessus. On divise la base de données en deux sous-ensembles : le premier sert à entraîner le système tandis que le second sert à évaluer des recommandations. L'algorithme simule une série de recommandations pour des utilisateurs

¹¹ Aujourd'hui, généreusement sponsorisée par de nombreuses entreprises comme Netflix, Facebook, Spotify, Zalando, Booking.com, pour ne citer qu'eux. La liste complète des sponsors de chaque édition peut être trouvée sur le site de la conférence (<https://recsys.acm.org/>).

enregistrés dans la base de données pour des objets dont l'évaluation réelle est connue dans le sous-ensemble d'évaluations. La précision de l'algorithme sera ensuite estimée sur base de la proximité entre la prédiction et le score réel. Plusieurs mesures d'erreur peuvent être utilisées mais c'est probablement la mesure de l'erreur quadratique moyenne qui monopolise l'attention (Konstan et Riedl, 2012). Pendant plusieurs années, la majorité des recherches réalisées dans ce domaine visera à minimiser cette erreur dans des conditions expérimentales bien déterminées (Seaver, 2019a).

Ce paradigme culminera en 2009 quand Netflix accordera, au bout d'une compétition de plusieurs années qui aura mobilisé 20.000 équipes issues de plus de 150 pays, un prix d'un million de dollars à une équipe de chercheurs pour être parvenus à développer un système de recommandation collaboratif dont l'erreur quadratique moyenne est 10 % plus faible que celle de Cinematch, l'algorithme à l'époque en service chez Netflix¹². L'algorithme lauréat ne sera jamais utilisé par le commanditaire de la compétition. Surspécialisé pour réduire l'erreur de prédiction, il est complexe, difficile à utiliser et beaucoup trop cher en ressources de calculs. Mais surtout, constate Seaver (2019b), il ne correspond plus aux besoins de l'entreprise.

De la surcharge informationnelle à la captation de l'attention

Les débuts de la recherche dans le domaine de la recommandation automatique sont marqués par un souci prononcé pour les enjeux de surcharge informationnelle. A ce moment, ce domaine est encore très proche de celui de la Recherche d'Informations. Lorsque la compétition de Netflix démarre en 2006, l'entreprise est encore principalement un service de location de DVDs et commence seulement à développer son service de streaming. Ses besoins en termes de recommandation sont très spécifiques. En 2008, le site s'apparente alors plus à un magasin en ligne et le travail réalisé sur l'expérience utilisateur vise à garantir l'efficacité et la rapidité de l'achat (Lindberg, 2019). Si la recommandation est déjà bien présente, elle cohabite avec beaucoup d'autres fonctionnalités. Elle est alors beaucoup plus limitée et se concentre sur l'identification de genres similaires à l'historique de consultation de l'utilisateur.

¹² Les performances de l'algorithme furent mesurées sur un jeu de données mis à disposition par Netflix pour toute la durée de la compétition. Pour plus de détails au sujet de cette compétition, nous renvoyons le lecteur vers Amatriain et Basilico (2015) mais également vers Hallinan et Striphas (2016).

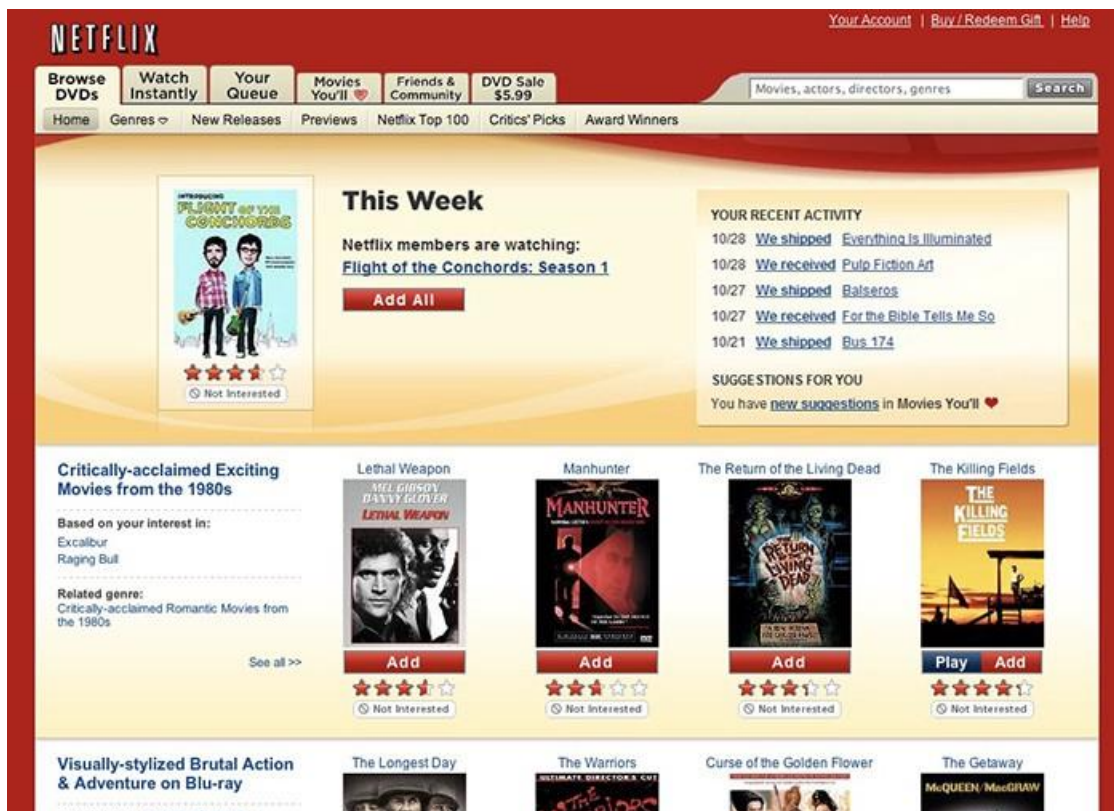


Figure 5 — Capture d'écran de l'interface Netflix en 2008 issu de Lindberg (2019)

Au fil des années, avec l'évolution des technologies et la popularisation du streaming, la recommandation automatisée investit de plus en plus l'interface de la plateforme. Chaque dimension de l'activité de recherche d'un film devient sujette à une recommandation. Dans un article publié en 2016, des employés de l'entreprise détaillent les principaux systèmes de recommandation intégrés dans la plateforme (Gomez-Urbe et Hunt, 2016). La structure de la page d'accueil repose sur une représentation matricielle comprenant jusqu'à 40 rangées et 75 vidéos par rangée. Pour chaque rangée, un algorithme organise les contenus du plus pertinent à gauche au moins pertinent à droite en fonction du genre de la rangée et du profil de l'utilisateur. Un autre algorithme s'occupe de définir l'ordre des genres à afficher verticalement. Une rangée spécifique dispose de son propre algorithme pour définir les contenus les plus pertinents sur l'ensemble du catalogue, indépendamment de leur genre. Une autre rangée consacrée aux contenus populaires dispose également de son propre système.



Figure 6 — Deux exemples d’une page Netflix incluant plusieurs rangées de recommandations issues de Gomez-Urbe et Hunt (2016)

Au-delà de la page d’accueil, plusieurs autres algorithmes de recommandation vont également structurer la navigation de l’utilisateur. Si les résultats d’une recherche sont également ordonnancés selon le profil de l’utilisateur, la fiche descriptive qui accompagne chaque média du catalogue sera également sujette à être personnalisée. Autant la vignette illustrant le média que les métadonnées composant la description seront soigneusement sélectionnées pour offrir à l’utilisateur les informations qui seront considérées comme les plus pertinentes afin d’aiguiller son choix (équipe de réalisation, récompenses, synopsis, acteurs, etc.). Cette fiche descriptive sera également accompagnée d’une suggestion de films ou de séries similaires au contenu sélectionné. Enfin, le contenu qui sera suggéré à l’utilisateur à la fin de son visionnage sera également personnalisé selon son profil de navigation. Au total, 8 algorithmes de recommandation différents cadrent l’expérience utilisateur sur Netflix.

L’ensemble de l’écran devient une mosaïque dynamique qui se réarrange en fonction des comportements de l’utilisateur. Ce glissement vers une recommandation omniprésente s’accompagne d’un impératif général de transparence. L’interface doit s’effacer pour garantir un accès presque « immédiat » à la collection médiatique. Cette transparence fonctionnelle ne doit pas être confondue avec l’impératif normatif de transparence que l’on retrouve dans de nombreuses chartes éthiques portant sur le traitement algorithmique de données personnelles. Dans les faits, les deux dynamiques s’opposent l’une l’autre. Garantir une expérience utilisateur « transparente »

implique généralement d'opacifier les mécanismes du système pour assurer la fluidité de l'action là où les débats éthiques actuels portent sur la réintégration d'une friction dans l'interaction¹³.

Seaver (2019a) constate que ce pivotement paradigmatique n'est pas seulement technologique mais également théorique. Au début des années 2010, l'industrie numérique, principalement dans la Silicon Valley, connaît un intérêt notable pour une littérature d'inspiration behavioriste reposant notamment sur les recherches de Skinner (voir par exemple Ferster et Skinner (1957)). Fogg incarne la figure de proue de ce courant à travers ses recherches au Persuasive Technology Lab de Stanford qu'il fonde en 1998. Fogg consacre dès lors ses recherches à la compréhension des leviers d'actions non-coercitifs permettant de provoquer un changement dans les attitudes et les comportements (voir par exemple Fogg, 2003). Ce courant irriguera une multitude d'ouvrages et de formations à destination d'entrepreneurs souhaitant capitaliser sur les principes de ce *behavioral design* pour assurer le conditionnement de leurs utilisateurs (Seaver, 2019a). Comme l'observe très justement Seaver (2019a), cette littérature entrepreneuriale s'inspire très largement des travaux de Skinner sur le renforcement positif :

Caged birds, given a lever that released food, would learn to press it; if experimenters adjusted the lever to only release food intermittently, the pigeons would learn to press it incessantly. Replace the pigeons with people, build the right levers into your product, and you too might amass a user base of compulsive lever-pressers. (p.4)

Petit à petit, ce courant s'est émancipé de ses têtes pensantes pour devenir une partie intégrante de l'imaginaire professionnel de l'industrie informatique. La figure de l'utilisateur ne repose alors plus sur le modèle classique de l'homo economicus rationnel mais plutôt sur celui de l'individu irrationnel, victime de ses biais cognitifs.

Avec ce changement de paradigme, c'est une toute autre façon d'évaluer la performance du système qui émerge. Recourir à la mesure de la précision de l'algorithme comme proxy de la satisfaction de l'utilisateur est de plus en plus critiqué (Knijnenburg *et al.*, 2011 ; McNee *et al.*, 2006 ; Pu *et al.*, 2011). Dans un article intitulé « Being accurate is not enough », McNee *et al.* (2006) plaident pour le développement de nouvelles métriques d'évaluation plus à l'écoute des besoins des

¹³ Nous reviendrons plus en détails sur cette problématique de la transparence dans le deuxième chapitre.

utilisateurs. Petit à petit, la mesure de l'erreur quadratique moyenne se fait détrôner par ce que Seaver (2019a) appelle des *métriques de captivation*. Ce tournant behavioriste encourage l'industrie à se détourner des cotations explicitées pour privilégier les signaux implicites qui sont considérés comme plus représentatifs des intentions réelles de l'utilisateur. Si le recours à ces signaux implicites n'est pas neuf, il prend une toute autre dimension grâce au développement du *streaming* et à la génération massive de logs à chaque interaction de l'utilisateur avec le système (Seaver, 2019a). La satisfaction de l'utilisateur est alors mesurée par la multiplication de signaux indiquant l'usage répété d'une fonctionnalité ou la consultation fréquente d'un contenu. Le plus important devient donc la durée de l'engagement de l'utilisateur avec le système. Au sujet de Netflix, Gomez-Urbe et Hunt (2016) décriront le rôle de la personnalisation pour leur entreprise en ces termes :

Our recommender system helps us win moments of truth: when a member starts a session and we help that member find something engaging within a few seconds, preventing abandonment of our service for an alternative entertainment option [...] Personalization allows us to significantly increase our chances of success when offering recommendations [...] But, most important, when produced and used correctly, recommendations lead to meaningful increases in overall engagement with the product (e.g., streaming hours) and lower subscription cancellations rates. [...] Over years of development of personalization and recommendations, we have reduced [our subscriber monthly] churn by several percentage points. Reduction of monthly churn both increases the lifetime value of an existing subscriber, and reduces the number of new subscribers we need to acquire to replace cancelled members. We think the combined effect of personalization and recommendations save us more than \$1B per year. (Gomez-Urbe et Hunt, 2016, p. 6-7)

Au fur et à mesure que ces technologies deviennent des atouts commerciaux, l'objectif principal n'est plus tant de faciliter l'exploration d'un catalogue mais de capter et préserver l'attention de l'utilisateur aussi longtemps que possible. Petit à petit, ce qui était auparavant un outil de navigation parmi d'autres devient progressivement un arrière-plan englobant la navigation de l'utilisateur.

Pouvoir algorithmique et cyberbalkanisation

En 1995, alors que les systèmes de recommandation n'en sont encore qu'à leurs balbutiements, Negroponte (1995) développe l'idée d'un journal entièrement personnalisé selon les intérêts de son lecteur et qu'il nomme *Daily Me* :

What if a newspaper company were willing to put its entire staff at your beck and call for one edition? It would mix headline news with « less important » stories relating to acquaintances, people you will see tomorrow, and places you are about to go to or have just come from. It would report on companies you know. In fact, under these conditions, you might be willing to pay the Boston Globe a lot more for ten pages than for a hundred pages, if you could be confident that it was delivering you the right subset of information. You would consume every bit (so to speak). Call it The Daily Me. (Negroponte, 1995, p. 153)

Cette hypothèse d'un accès médiatique personnalisé en fonction des préférences d'un individu suscite l'inquiétude de plusieurs chercheurs et essayistes qui s'interrogent sur les risques de fragmentation de l'espace public. Avec cet exemple, Negroponte illustre un aspect fondamental du pouvoir exercé par ces technologies d'ordonnancement de l'information : la gestion de la visibilité en ligne (Bucher, 2018). En fonction de la plateforme et du domaine d'application, l'algorithme détermine à la fois ce que l'utilisateur peut voir et comment il peut se faire voir. Une première formulation de cette inquiétude apparaît déjà en 1996. Dans un article intitulé « Electronic Communities: Global Village or Cyberbalkans », Van Alstyne et Brynjolfsson (1996) suggèrent que l'expansion des technologies de communication comporte le risque d'une balkanisation des communautés numériques. Là où McLuhan postule l'émergence d'une communauté unique avec sa notion de village planétaire, Van Alstyne et Brynjolfsson (1996) envisagent l'hypothèse inverse :

With the customized access and search capabilities of IT, individuals can focus their attention on career interests, music and entertainment that already match their defined profiles, or they can read only news and analysis that align with their preferences. Individuals empowered to screen out material that does not conform to their existing preferences may form virtual cliques, insulate themselves from opposing points of view, and reinforce their biases. Internet users can seek out interactions with like-minded individuals who have similar values, and thus become less likely to trust important

decisions to people whose values differ from their own. This voluntary balkanization and the loss of shared experiences and values may be harmful to the structure of democratic societies as well as decentralized organizations. (Van Alstyne et Brynjolfsson, 1996, p. 24)

Pour rappel, l'expression *système de recommandation* n'apparaît qu'en 1997. Lorsqu'ils parlent de « customized access and search capabilities of IT », les auteurs de cette critique font pourtant bien référence à des techniques de recommandation comme le témoigne une note de bas de page qui suit ce passage dans l'article original. Cette note de bas de page fait référence à *Firefly*, un système de recommandation musicale incluant un outil de filtrage collaboratif. Malgré cette critique, Van Alstyne et Brynjolfsson (1996) conservent une certaine nuance dans leur analyse qu'il nous semble importante de souligner. Ce risque de balkanisation n'est pas un mal absolu. Un certain niveau de spécialisation au sein d'un réseau peut permettre d'atteindre un niveau de stabilité et d'efficacité optimal. Toutefois, trop de ségrégation nuit à la collaboration et à l'innovation. L'analyse proposée par Van Alstyne et Brynjolfsson (1996) ne doit donc pas être lue comme un rejet de ces technologies mais comme un appel à une conception critique : « We can, and should, explicitly consider what we value as we shape the nature of our networks and infrastructure – with no illusions that a greater sense of community will inexorably result » (Van Alstyne et Brynjolfsson, 1996, p. 24).

Onze ans plus tard, Sunstein (2007) s'inscrit dans la continuité de cette critique en revenant explicitement sur le Daily Me de Negroponte. En plus des enjeux de collaboration et de circulation de l'information, Sunstein accentue particulièrement les risques de polarisation de l'opinion provoqués par une faible exposition à des opinions contraires. De plus, selon sa perspective, un écosystème médiatique équilibré doit aussi exposer ses usagers à un spectre d'expériences communes, ce qui n'est pas le cas avec ces technologies.

Enfin, la formulation la plus récente et la plus populaire de ces préoccupations découle de l'ouvrage publié en 2011 par Eli Pariser, intitulé *The Filter bubble - What the Internet is hiding from you* (Pariser, 2011). Dans les faits, la théorie de la bulle de filtres qu'il propose n'apporte pas vraiment d'éléments nouveaux qui la démarquent des travaux antérieurs. De plus, en versant occasionnellement dans un déterminisme assez tranché, il y perd la subtilité d'analyse de ses prédécesseurs. Le principal ajout de Pariser concerne l'attention qu'il porte à la qualité du contenu

recommandé. En prenant en compte la structure économique sur laquelle repose ces technologies, Pariser constate que cette recherche de l'attention pourrait conduire à privilégier des contenus de moindre valeur démocratique au détriment de contenus plus « importants » (en l'occurrence, l'actualité sociale et politique). Faisant un parallèle avec le sucre et l'obésité, il postule un risque de dégradation de la qualité du « régime informationnel » de l'utilisateur de ces services.

Depuis la formulation de ces critiques, un nombre important d'études empiriques se sont penchées sur l'effet de ces technologies sur les usages. Elles ont notamment contribué à nuancer ces critiques et à dessiner un portrait plus complexe des utilisateurs de ces services. Pour la section finale de ce chapitre, nous reviendrons sur ces études, afin de préciser les contours de la problématique générale de cette thèse.

DU POUVOIR ALGORITHMIQUE À SA MÉDIATION

Diversité d'algorithmes

Si la popularité de la théorie de la bulle de filtre a le mérite d'avoir attiré l'attention générale sur les enjeux et les effets de la personnalisation algorithmique, elle a pour faiblesse majeur de réifier la figure de l'algorithme sous la forme d'un objet aux contours figés. Or, la multiplication d'études focalisées sur des domaines d'applications spécifiques dessine une situation beaucoup plus complexe. Les lignes qui suivent sont consacrées à une revue de la littérature de ces recherches qui s'intéressent aux effets de la recommandation algorithmique sur l'exposition à une information diversifiée.

Si on a longtemps pensé que les résultats des recherches en ligne s'adaptent fortement au profil de l'utilisateur, des recherches plus récentes semblent indiquer que le chevauchement des résultats entre différentes requêtes d'utilisateurs différents reste important (Haim *et al.*, 2018 ; Krafft *et al.*, 2019 ; Puschmann, 2019). En ce qui concerne la recommandation d'actualités, selon une étude dirigée par Facebook, si l'algorithme de la plateforme occasionne bien une baisse de l'exposition à des opinions contraires, le choix des relations sociales des utilisateurs aurait une influence plus importante sur la personnalisation du flux d'actualités. Des études plus expérimentales semblent également attribuer un effet négatif sur la diversité des opinions rencontrées du fait de l'action d'algorithmes de recommandation. Claussen *et al.* (2019) observent que la personnalisation de la page d'accueil d'un site d'actualité diminue la diversité des contenus consultés tandis que Dylko

et al. (2017) constatent que la personnalisation d'actualités limite l'exposition à des informations dissonantes.

Toutefois, des études portant sur d'autres domaines d'application viennent nuancer ces observations. Aiello et Barbieri (2017) se sont intéressés à la recommandation de contacts sur Flickr et Tumblr. Ils observent que les recommandations automatisées accroissent la diversité du réseau social des utilisateurs qui y sont confrontés. Selon les auteurs de l'étude, le système expose l'utilisateur à une plus grande variété de contacts que si ce dernier avait exploré les blogs ou fait des recherches directement sur le site.

Dans le cas des services de streaming vidéo, l'étude de Roth *et al.* (2020) montre que la recommandation de vidéos sur Youtube dirige l'utilisateur vers des groupes de vidéos homogènes privilégiant des vidéos populaires. Ces observations s'alignent sur les résultats de Ledwich et Zaitsev (2019). En se penchant sur le site MovieLens qui recommande des films sur base des scores attribués par la communauté d'utilisateurs, Nguyen *et al.* (2014) constatent que les utilisateurs qui ont recours à un algorithme de recommandation consultent des contenus de moins en moins diversifiés au fil du temps. Toutefois, cette baisse est moins importante que chez les autres utilisateurs de la plateforme qui n'utilisent pas les recommandations. Du côté de Spotify, la consommation musicale guidée par l'usage de systèmes de recommandation serait associée à une diminution de la diversité sur le long terme (Anderson *et al.*, 2020). Toutefois, Holtz *et al.* (2020) ainsi que Datta *et al.* (2017) observent un effet positif de la recommandation automatisée sur la diversité des pratiques d'écoute. Holtz *et al.* (2020) constatent néanmoins que, même si plus de contenus différents sont consultés à l'échelle de la plateforme, chaque individu a tendance à se retrancher derrière des intérêts plus spécifiques. Enfin, les travaux de Beuscart *et al.* (2019) sur Spotify indiquent que la recommandation automatique contribue à guider l'utilisateur vers des titres moins populaires et s'associe à la démarche exploratoire d'utilisateurs éclairés.

Il faut souligner que les études citées ici sont très inégales dans la portée de leurs résultats, la nature de leur méthodologie et la taille de leurs échantillons, ce qui complique la comparaison de leurs résultats. En l'état actuel des choses, il n'est pas possible de réduire l'effet des algorithmes de recommandation sur les usages informationnels à un isolement progressif dans des espaces homogènes. Si l'intégration de systèmes de recommandation a une influence non négligeable sur

l'horizon informationnel des individus, elle n'est toutefois pas aussi uniforme que pourrait le laisser penser la théorie des bulles de filtres et ne concerne qu'une minorité d'utilisateurs.

Enfin, nous pouvons légitimement nous demander dans quelle mesure se concentrer uniquement sur la diversité des recommandations est une focale d'analyse pertinente pour évaluer l'effet de ces systèmes. Comme le souligne Rieder (2020):

What constitutes a 'good' recommendation in the context of a music service may produce troubling effects when applied to politically salient content: the same capacity to drill deeper and deeper into specialized genres may yield hidden gems on Spotify but lay pathways to increasingly radicalized political positions on YouTube. Content filtering optimized for user engagement may produce benign results in one kind of context but stir violence in another. (p.334)

En fonction du domaine, il est fort probable que l'usage de systèmes de recommandation puisse autant contribuer à un enfermement idéologique de certains utilisateurs qu'à leur *empowerment* (Zuiderveen Borgesius *et al.*, 2016).

Diversité des pratiques

Les préoccupations relatives au développement de la recommandation automatisée reposent également sur le postulat d'une généralisation de cette logique d'accès à l'information ainsi que sur une représentation passive de l'utilisateur. Une littérature de plus en plus complète sur les usages numériques nous permet de nuancer ces deux affirmations.

A l'heure actuelle, plusieurs travaux semblent indiquer que les pratiques médiatiques restent caractérisées par une diversité des modalités d'accès à l'information. Fletcher et Nielsen (2017) observent que dans le contexte d'un écosystème informationnel avec une offre médiatique diversifiée, la plupart des usagers combinent des contenus issus de médias de niche avec d'autres issus de médias s'adressant à une population plus large. Ces résultats corroborent les conclusions de Dubois et Blank (2018) qui mettent en évidence qu'un utilisateur de réseaux socio-numériques se limite rarement à une seule source d'information et qu'il tend à croiser les résultats de ses recherches (recherches en ligne, consultation de médias dits « traditionnels », de médias sociaux, etc.). Resnick *et al.* (2013) constatent également que même l'usage de sites d'actualité partisans s'accompagne généralement d'une consommation d'actualités issues de canaux plus généraux et

consensuels. Enfin, contrairement à ce que l'on pourrait croire, les médias traditionnels restent encore largement consultés selon Newman *et al.* (2020). Il est dès lors assez rare qu'un utilisateur de plateformes en ligne se retrouve dans un environnement numérique parfaitement clos et ne répondant qu'à une logique algorithmique dominante.

Souligner cette complexité des usages ne signifie pas que l'utilisation des algorithmes de recommandation n'est pas sans conséquence, mais que l'analyse de leurs effets doit se faire au prisme des contextes d'usage afin de dépasser une représentation déterministe de l'utilisateur qui se retrouve dépouillé de toute agentivité.

Lorsque les usages sont observés depuis la réalité quotidienne des internautes, l'emprise des calculateurs sur leur vie semble s'évaporer. Les usages sont beaucoup plus vagabonds, diversifiés et stratégiques que ne le pensent ceux qui raisonnent depuis une seule plateforme. [...] Il est des moments où les internautes choisissent le confort du guidage et d'autres où ils débrayent pour explorer et se perdre. La réduction de leurs pratiques à des automatismes comportementaux fait oublier que les usages d'Internet ne cessent de se complexifier, de s'intellectualiser et de devenir en eux-mêmes des objets réflexifs. (Cardon, 2015, p. 103)

Si nous nous focalisons plus spécifiquement sur l'usage de systèmes algorithmiques, plusieurs travaux remettent également en question cette image de l'utilisateur passif qui suit naïvement le fil de ses recommandations en interrogeant les attitudes et les affects que génèrent ces systèmes chez leurs utilisateurs.

Comme le constate Bucher (2017), les recherches sur la perception des algorithmes sont assez peu nombreuses et leurs résultats encore très contradictoires. A travers une étude qualitative comprenant 40 utilisateurs de Facebook, Eslami *et al.* (2015) constatent que plus de la moitié des participants ne sont pas conscients de la présence d'un algorithme qui régule leur fil d'actualités. A l'inverse, 75 % des répondants de l'étude par questionnaire menée par Rader et Gray (2015) déclarent qu'ils sont conscients de ne pas avoir accès à l'ensemble des contenus disponibles. A travers une analyse qualitative de 25 utilisateurs de Facebook conscients de la présence d'un algorithme, Bucher (2017) s'est intéressée aux affects ressentis par ces individus lorsqu'ils sont confrontés à ces technologies. Bien que ses participants n'aient pas une connaissance précise des mécanismes à l'œuvre, ils développent toutefois des théories plus ou moins élaborées qui

alimentent leurs représentations mentales de l'algorithme. Bucher (2017) emploie le terme d'*imaginaire algorithmique* pour décrire ce faisceau de représentations. Elle constate notamment que cet imaginaire a des conséquences bien concrètes sur les attitudes et comportements des utilisateurs. Face à la menace d'invisibilité que fait peser la plateforme sur ses utilisateurs, ces derniers adaptent leur comportement pour être plus ou moins facilement « lus » par le système (Bucher, 2017). En complément de cette approche des affects et des imaginaires, une étude de Wiard *et al.* (2022) s'est intéressée aux pratiques concrètes mises en œuvre par des jeunes utilisateurs de plateformes numériques. Même si les participants de leur étude confirment ne pas toujours avoir les mots pour faire sens de leur environnement numérique, ils parviennent malgré tout à moduler leurs interactions avec l'algorithme, en choisissant quand interagir et quand s'abstenir, de sorte à orienter le système selon leurs intérêts.

En privilégiant les comportements passés comme prédicteurs de ses intérêts futurs, le système renvoie, par défaut, l'utilisateur à un double numérique qui n'est qu'un reflet comportemental aussi partiel que partial. En même temps qu'il développe son imaginaire algorithmique, l'utilisateur se confronte à sa doublure numérique. A travers les catégories que le système projette sur lui, qu'elles soient correctes ou non, l'algorithme pénètre le processus de construction identitaire de l'individu en demandant à l'utilisateur de se positionner par rapport aux étiquettes qui lui sont apposées (de Vries, 2010). Selon Bucher (2017), en complément des enjeux d'accès à l'information, la compréhension des conditions de formation de ces imaginaires est fondamental à l'analyse et à la compréhension du pouvoir social de ces technologies.

Médiation de l'algorithme

Les études citées dans les deux sections précédentes contribuent à mettre en évidence toute la difficulté de production d'un savoir général portant sur les effets des systèmes de recommandation sur les usages. Ceci s'explique notamment par la variété des domaines d'application, des contextes d'usage, des pratiques de consultation et, surtout, des technologies employées. L'historique que nous avons présenté précédemment pourrait donner l'illusion d'une évolution fluide et linéaire de ces technologies. En réalité, nous n'avons fait que mettre en lumière un seul fil d'une toile beaucoup plus complexe qui combine plusieurs disciplines comme la recherche d'information, le *machine learning* ou encore la théorie des graphes. Au fil du temps, chacune de ces disciplines a

contribué à l'élaboration de nouveaux blocs élémentaires qui viennent s'ajouter aux précédents pour constituer le domaine des solutions à disposition des architectes de ces systèmes :

A document search engine may build on manually created indexes, implement collaborative tagging, or extract word frequencies from text contents, taking into account document structure, word co-occurrence, or formatting. It could use a term-weighting scheme like tf-idf to add discriminatory power. It could model a query language using Boolean or more intricate parameters, or it could sway into another direction and attempt to map the document space using something like PageRank or even a clustering technique that provides an overview to users rather than a ranking. The system could incorporate a feedback loop, use personalization, take into account location, device type, or time of day as means to differentiate. And each one of these things could be done in many different ways. To recognize these spaces of variation and choice is startling, because it means that technologies like computing or even algorithmic information ordering cannot be totalized into a singular logic, as convenient as that would be. (Rieder, 2020, p. 331)

Au-delà de ce premier constat, cette revue de littérature met également en évidence le fossé qui s'est creusé entre l'importance accordée à l'algorithme de celle accordée à sa médiation par l'interface du système. Pourtant, c'est bien l'interface qui définit la nature de la relation qu'il est possible d'établir avec l'algorithme en modulant ce qu'il est possible de savoir à son sujet et en déterminant les modalités d'actions sur son comportement. Elle est au cœur de la relation de pouvoir qui lie l'utilisateur aux intentions des concepteurs du système.

Si l'expérience utilisateur et l'interfaçage de l'utilisateur avec l'algorithme sont des préoccupations très présentes dans le domaine de l'interaction homme-machine, les recherches issues de ce domaine se limitent souvent à la mesure de la satisfaction de l'utilisateur afin d'améliorer l'utilisabilité et la précision de l'algorithme de recommandation. Toutefois, un nombre croissant d'études issues des sciences sociales empruntent également cette trajectoire en élargissant la focale de l'analyse. Drumond *et al.* (2018) se sont ainsi intéressés à la façon dont l'interface de Netflix, à travers le recours à certains éléments de langage, encourage l'adoption de « bons usages » et le partage de ses données personnelles pour affiner les recommandations. A l'inverse, Karakayali *et al.* (2018) se sont intéressés à l'interface d'une plateforme de streaming musical pour comprendre

ce qui conduisait certains utilisateurs à placer l'algorithme au centre d'une pratique de raffinement de leur identité musicale.

Les travaux de Bucher (2017) et de Wiard *et al.* (2022) nous indiquent que les relations de pouvoir qui s'instaurent pour l'instant entre les utilisateurs de plateformes numériques et leurs infrastructures techniques sont encore assez fluides pour permettre le développement de tactiques de contournement. Toutefois, la tendance actuelle en terme de conception semble conduire à une rigidification de ces relations comme peut le témoigner l'interface d'un site comme Tiktok dont la zone d'affichage se réduit presque uniquement à ses recommandations. Dit autrement, selon les termes de Davis (2020), ce qui était auparavant du registre du *découragement* évolue progressivement vers le registre de l'*exigence*. La suite de cette thèse s'inscrit dans cette problématique en cherchant à mieux comprendre comment certaines modalités d'interaction homme-machine influent sur la perception de l'algorithme ainsi que sur son appropriation critique.

Toutefois, contrairement à ce qui se fait encore très souvent dans le domaine des sciences sociales, nous souhaitons inscrire notre critique dans une démarche proactive plutôt que réactive. Nous reprenons ici l'appel de Rieder (2020) pour le développement, dans la lignée de Simondon, d'une réelle *mécanologie* qui ne se limite pas à corriger la technique mais nous invite à questionner plus fondamentalement les formes alternatives qu'elle peut incarner. Trop souvent encore, la critique se limite à un registre défensif, mobilisant le vocabulaire de la transparence, de l'*accountability*, de la vie privée, de l'auditabilité. Plutôt que de se restreindre à l'analyse des insuffisances et des biais, il est nécessaire de développer des formes de contributions scientifiques qui reposent sur l'expérimentation et la création. A travers cette critique, Rieder appelle à la reconquête d'une forme d'utopisme technologique qui évite le piège d'un solutionnisme naïf mais qui refuse tout autant un conservatisme obstiné.

Appliqué au domaine des systèmes de recommandation, cette approche demande de dépasser la lorgnette de la diversité des recommandations pour élargir l'analyse aux modes de fonctionnements alternatifs sur lesquels baser ces dispositifs. Pouvons-nous envisager une configuration qui ne cherche pas à réduire les possibilités d'appropriation mais qui vienne les développer ? Plutôt que de glisser vers l'exigence d'un assujettissement à l'algorithme, serait-il possible de glisser vers le registre de l'encouragement d'un usage critique ? Dès lors, quelle forme doit prendre cet usage et quelles modalités d'interaction sont les plus à même de le stimuler ? Pour reprendre les mots de

Seaver qui utilise la métaphore du piège (*trap*) pour désigner les algorithmes de recommandation :
« The question to ask of traps may not be how to escape from them, but rather how to recapture them and turn them to new ends in the service of new worlds. » (Seaver, 2019a, p. 13).

CHAPITRE 2 — INTERFACES ET AUTONOMIE

Bien que Sunstein (2007) retienne surtout le *Daily Me* de Negroponte (1995), celui-ci parle également d'un *Daily Us* qui serait le complément communautaire du *Daily Me*. Au sujet de l'interaction entre ces deux modèles, Negroponte dit ceci :

These are not two distinct states of being, black and white. We tend to move between them, and, depending on time available, time of day, and our mood, we will want lesser or greater degrees of personalization. Imagine a computer display of news stories with a knob that, like a volume control, allows you to crank personalization up or down. You could have many of these controls, including a slider that moves both literally and politically from left to right to modify stories about public affairs. (Negroponte, 1995, p. 154)

Negroponte pose ici les bases de ce qui sera appelé plus tard, dans le vocabulaire de l'interaction humain-machine, la contrôlabilité de l'algorithme. Cette contrôlabilité désigne l'ensemble des éléments de l'interface qui permettent à l'utilisateur d'agir sur le fonctionnement du système. Mais est-ce qu'il suffit de disposer de ces opportunités de contrôle pour les saisir ? N'est-ce pas contre-productif de vouloir contrôler un dispositif qui, par définition, implique de lui déléguer une forme d'autorité ? La question du pouvoir algorithmique que nous avons discutée dans le chapitre précédent pose, en creux, celle de l'autonomie du sujet et de sa capacité à garder le contrôle sur son environnement. Qu'est-ce qu'être libre ou autonome lorsque l'on utilise un système de recommandation ? C'est à travers le prisme de la notion d'algorithme de service public que nous proposons d'interroger ces enjeux tout au long de ce deuxième chapitre.

Cette problématique de l'autonomie ouvre un domaine d'investigation particulièrement large qui couvre plusieurs disciplines et plusieurs notions connexes telles que la liberté ou l'agentivité. Une revue de littérature exhaustive des travaux relatifs à ce questionnement est une entreprise qui dépasse le cadre de cette thèse. Notre objectif, plus modeste, consistera principalement à identifier les principales interprétations de cette notion au sein des travaux qui se consacrent spécifiquement aux systèmes de recommandation et à identifier leurs origines théoriques. Nous nous attarderons plus particulièrement sur l'interprétation de l'autonomie comme « réalisation de soi » proposée par

Knijnenburg (2016) pour sa popularité dans le domaine de l'interaction humain-machine et des systèmes de recommandation.

Ce cadrage comporte toutefois plusieurs défauts sur lesquels nous reviendrons en détail et qui nous empêcheront de mobiliser cette approche dans le cadre de notre projet de recherche. Pour compenser ces limites, nous passerons par les travaux de Foucault sur le sujet, mais également par plusieurs travaux issus de la philosophie morale pour arriver à une formalisation de l'autonomie qui puisse rendre compte de la nature distribuée de notre agentivité et qui puisse être traduite en principes de design concrets.

ALGORITHME DE SERVICE PUBLIC

Cadrage général

La notion d'*algorithme public* ou d'*algorithme de service public* n'a pas d'origine clairement identifiable. La notion apparaît notamment dans un guide institutionnel publié par Etalab¹⁴ et qui attribue à ces algorithmes 3 caractéristiques principales : ils opèrent au service de l'intérêt général, ils exécutent le droit et sont souvent incontournables (c.-à-d. qu'il n'existe pas d'alternatives pour le citoyen). Toujours selon ce guide :

Les algorithmes publics sont des formes de l'action publique et sont à ce titre soumis à la même forme d'exigence de redevabilité. Les administrations qui conçoivent et utilisent des algorithmes publics doivent donc « rendre des comptes » de leur utilisation auprès des individus concernés, mais aussi de la société dans son ensemble. (Etalab, 2022, p. 6)

La notion apparaît également dans un sens plus restreint pour désigner les algorithmes employés par des médias de service public. La philosophie reste très similaire et vise à promouvoir le développement d'outils automatisés reposant sur des logiques différentes des logiques de marché habituelles (Helberger, 2015). Dans un article publié en 2018, le responsable data de la BBC défend le développement d'algorithmes de recommandations dont les principes généraux s'alignent sur les valeurs de service public (Fields *et al.*, 2018). A ce moment-là, il parle notamment des 3 fameuses missions de la BBC énoncées par John Reith, son premier directeur général : informer, éduquer, divertir. Il identifie ainsi deux objectifs principaux de développement : garantir la

¹⁴ Administration publique française chargée de superviser la gestion par l'Etat français de données numériques

diversité des recommandations ainsi qu'assurer la capacité de ces services à rendre des comptes (principe d'*accountability* que nous traduirons par responsabilité, dans la suite de ce document). Lorsque le directeur du numérique et de la production de Radio France, Laurent Frisch, parle de leur algorithme de service public, il souligne que celui-ci est « pensé pour susciter la surprise et la découverte » plutôt que de garantir l'engagement des utilisateurs de services commerciaux (Schmitt et Madelaine, 2019, para. 7).

Dans son bilan du contrat de gestion de la RTBF 2019-2022, le CSA (2022) définit un algorithme de service public selon 4 missions :

Suggérer aux utilisateurs une offre généraliste et diversifiée, ou autrement dit offrir aux publics non seulement ce qu'ils souhaitent voir, mais élargir leurs horizons en leur présentant des contenus auxquels ils ne seraient autrement pas exposés ; Eviter la création de « bulles de filtres » ou « bulles cognitives » ; Agir en complémentarité de l'éditorialisation humaine, sans la remplacer ; Mettre en lumière le fonctionnement de l'algorithme dans une double démarche de transparence : à l'échelle du système de recommandation à travers une mesure de l'efficacité de la diversité des recommandations qui sont faites, et à l'échelle de chaque utilisateur en expliquant pourquoi un contenu a été recommandé. (p.33)

Il semble que les débats relatifs à cette notion d'algorithme de service public se stabilisent autour de 3 préoccupations : équité, responsabilité, transparence. Ces trois préoccupations sont assez communes dans le lexique des travaux en éthique de l'intelligence artificielle (Floridi *et al.*, 2018). Si dans ce domaine, la notion d'équité renvoie à la problématique d'un traitement « juste » et d'une absence de biais, elle prend toutefois une tournure quelque peu différente lorsqu'elle s'applique à notre problématique. Dans ce contexte, la notion d'équité renvoie plutôt à la concordance entre les logiques de l'algorithme et le respect d'une série d'impératifs de service public.

Éditorialisation algorithmique

Parmi ces 3 pôles, c'est cette question de la « ligne éditoriale » de l'algorithme, autrement dit de la réconciliation entre les valeurs de service public et les opportunités technologiques, qui capte la majorité des préoccupations (Álvarez *et al.*, 2020 ; Fields *et al.*, 2018 ; Sørensen, 2019). Le « caractère » final de l'algorithme de recommandations (sa propension à faire des recommandations

précises, diversifiées, aléatoires etc.) dépend en grande partie de la nature de l'évaluation à laquelle il sera soumis lors du paramétrage du dispositif. Obtenir une recommandation plus « éthique » ou plus « juste » implique donc de définir formellement ce à quoi correspond cette éthique ou cette justice de sorte à pouvoir établir un protocole d'évaluation permettant de juger de la proximité des recommandations à l'idéal recherché. Toutefois, il existe assez peu de chevauchement entre les mesures d'évaluation habituelles d'un système de recommandation et des principes de service public. Grün et Neufeld (2021) identifient deux logiques traditionnelles d'évaluation:

- Métriques traditionnelles issues du *machine learning* et du domaine de la recherche d'information : *recall*, précision etc. Ces métriques sont principalement issues du monde académique et sont utilisées pour mesurer la qualité des prédictions sur base d'un jeu de données connues. Ce sont donc des mesures dites *offlines* (qui sont faites avant mise en ligne du service)
- *Key Performance Indicators* (KPIs) issus des secteurs industriels : nombre de minutes passées sur un contenu, nombre de clics etc. Ces métriques sont utilisées pour mesurer l'engagement des utilisateurs et indirectement, leur satisfaction. Ces mesures sont calculées quand le service est en ligne.

Un nombre important de recherches, souvent réalisées en partenariat avec un média de service public, visent à formaliser les valeurs de service public, à les traduire en termes techniques et à identifier les méthodes de combinaisons les plus adéquates avec les mesures plus traditionnelles (Álvarez *et al.*, 2020 ; Helberger *et al.*, 2016 ; Smets *et al.*, 2020 ; Sørensen, 2019 ; Tallerås *et al.*, 2020). La valeur la plus récurrente dans ces débats concerne la préservation d'une certaine diversité de l'information diffusée. Cette focale provient notamment de la popularité des critiques formulées par Eli Pariser et ses prédécesseurs qui furent discutées dans le premier chapitre. Deux difficultés principales demeurent.

Dans un premier temps, resserrer la problématique de l'éditorialisation algorithmique à des enjeux de diversité ne suffit pas à faciliter l'exercice de formalisation. La littérature technique fait généralement la distinction entre diversité (Knijnenburg, Willemsen, *et al.*, 2012), nouveauté (Castells *et al.*, 2015 ; Vargas et Castells, 2011), sérendipité (Kotkov *et al.*, 2018) et couverture (*coverage*) (Vrijenhoek *et al.*, 2021). Dans ce contexte, la diversité d'un set de recommandation s'exprime par la différence entre les éléments de cette liste. Cette différence peut être estimée de

différentes façons en mesurant une distance lexicale ou thématique par exemple. La nouveauté d'une recommandation désigne son écart par rapport à ce qui est habituellement recommandé tandis que la sérendipité qualifie la propension d'une recommandation à surprendre positivement l'utilisateur. Enfin, la couverture évalue le pourcentage de recommandations potentielles qu'un système peut émettre en regard de ce qui est disponible dans la base de données. Parallèlement à ces formalisations techniques de la notion de diversité, Helberger (2019) identifie au moins 4 conceptions normatives de la diversité qui se raccrochent à des théories démocratiques différentes :

- Le modèle libéral : Promouvoir le développement personnel et la prise de décision autonome. Le système doit donc permettre à l'utilisateur de se spécialiser dans un domaine d'intérêt.
- Le modèle participatif : Faciliter l'engagement citoyen au sein de la société à travers la recommandation de thématiques identifiées comme importantes à connaître tout en garantissant une représentation proportionnelle de tous les points de vue
- Le modèle délibératif : Faciliter le développement d'un avis collectif à travers la présentation des débats contemporains et d'une variété d'opinions
- Le modèle critique : Rendre visible des opinions issues de groupes marginaux

En plus de cette richesse conceptuelle, une autre difficulté majeure concerne l'équilibre à adopter entre mesures traditionnelles et impératifs normatifs. A quel moment est-ce que privilégier une certaine diversité au détriment de la précision de la recommandation devient plus néfaste que de ne pas privilégier cette diversité (Fields *et al.*, 2018) ?

De l'éducation à la transparence

Nous avons déjà souligné, dans le premier chapitre, la prédominance de la figure de l'algorithme au sein du paysage scientifique des systèmes de recommandation. Cette prédominance se voit également au sein de la littérature technique qui relègue souvent l'expérience de l'utilisateur au second rang des préoccupations (McNee *et al.*, 2006). Il n'est dès lors pas surprenant de faire le même constat dans l'extension de ce domaine appliqué aux médias de service public. Cette situation peut aussi s'expliquer par le statut principal du média de service public comme diffuseur de contenus. Sous cette perspective, la majorité des préoccupations tournent autour de la logique

éditoriale qui doit se distinguer des offres commerciales classiques répondant aux principes de marché.

Toutefois, il nous semble que la notion de service public s'est diversifiée au fil du temps pour introduire de nouvelles missions. En effet, l'investissement croissant de ces médias de service public dans des activités d'éducation aux médias, qu'il soit volontaire ou imposé par une autorité subsidiaire, traduit une perception du service public comme vecteur de développement d'un rapport critique à l'information. Ainsi, dans un rapport de 2017, le Conseil de l'Europe (2007) stipule que :

Public service media should continue to play a central role in education, media literacy and life-long learning, and should actively contribute to the formation of knowledge-based society. Public service media should pursue this task, taking full advantage of the new opportunities and including all social groups and generations. (section 1.e.21)

Plus récemment, le 5ème contrat de gestion de la RTBF (Cinquième contrat de gestion consolidé de la RTBF incluant son premier avenant, 2020) comprend l'extrait suivant :

La RTBF considère que l'éducation, dans un monde où les repères fluctuent sans cesse, est un des moyens de le décoder et que l'accès à la connaissance, à la compréhension des enjeux et à leur évaluation avec un regard critique participe à son rôle de service public dans un débat démocratique animé sur toutes les plateformes des nouvelles technologies de communication numériques. (p.3)

Il nous apparaît dès lors évident que l'interface des dispositifs numériques développés par ces organismes se doit de refléter ces engagements en facilitant le développement d'un rapport critique aux algorithmes. A l'heure actuelle, cette dimension du service public est principalement traitée sous l'angle de la « transparence » du système. A travers son *Digital Services Act*, la Commission Européenne espère ainsi imposer une obligation de transparence à tous les systèmes de recommandations mis en place par des plateformes numériques. Derrière cette idée de transparence, on retrouve deux dynamiques différentes avec d'une part une obligation de responsabilité, qui passe souvent par la publicisation des principes généraux de l'algorithme, et une obligation de transparence à l'égard de l'utilisateur. Selon l'article 24a de l'accord provisoire du *Digital Services Act* du 23 avril 2022 (Regulation 2022 of the European Parliament and of the

Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, 2022) :

1.Providers of online platforms that use recommender systems shall set out in their terms and conditions, in plain and intelligible language, the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters.

2.The main parameters referred to in paragraph 1 shall explain why certain information is suggested to the recipient of the service. They shall include, at least:

- a) the criteria which are most significant in determining the information suggested to the recipient of the service;
- b) the reasons for the relative importance of those parameters.

3.Where several options are available pursuant to paragraph 1 for recommender systems that determine the relative order of information presented to recipients of the service, providers of online platforms shall also make directly and easily accessible from the specific section of the online platform's online interface where the information is being prioritised a functionality allowing the recipient of the service to select and to modify at any time their preferred option. (Article 24a)

Nous pouvons constater que dans ce texte, l'obligation de transparence à l'égard de l'utilisateur concerne surtout la présence d'explications sur le fonctionnement du système ainsi que sur les options permettant à l'utilisateur de modifier ce fonctionnement. Permettre à l'utilisateur de contrôler l'algorithme n'est pas rendu obligatoire mais, quand de telles fonctionnalités existent, elles doivent être clairement indiquées.

Il serait absurde de dire que l'ensemble de la recherche sur le sujet de la recommandation de service public se concentre uniquement sur l'algorithme. Au contraire, certaines contributions n'hésitent pas à appréhender l'ensemble du système (Bernstein *et al.*, 2021 ; voir par exemple : Harambam *et al.*, 2018). Toutefois, dans l'ensemble, ces démarches sont isolées et ne partagent que peu de terrain commun contrairement aux recherches orientées sur l'algorithme en lui-même qui profitent d'une plus grande intersectionnalité. Il nous semble que ce constat peut être expliqué par l'absence d'une problématisation claire des enjeux de conception de ces interfaces.

Médiation de service public

La faiblesse de ce modèle équité/responsabilité/transparence réside dans le travail de spécification insuffisant du pôle « transparence » et sa caractérisation comme une finalité et non comme un ensemble de moyens. Cette idée de la transparence comme une finalité s'accroche à une approche cognitiviste de la connaissance au sein de laquelle la communication se retrouve désincarnée. Dépouillée d'enjeux de médiation, elle en devient une affaire de tuyauterie (Peraya et Meunier, 1999). Dès lors, rendre un algorithme transparent se limite à découper une fenêtre dans la boîte noire qui l'enserme pour permettre à tout un chacun de scruter sa mécanique. Il ne reste alors plus qu'à apprendre la programmation et la « pensée informatique » (Wing, 2011) pour permettre le développement d'un jugement critique sur le système qui leur est présenté. Nous noterons au passage que les analyses issues des domaines plus techniques, comme l'interaction homme-machine, sont souvent construites sur une représentation d'un utilisateur purement rationnel à qui il ne manque que l'information adéquate pour pouvoir agir sur le système, sans considération pour les disparités possibles d'une communauté d'utilisateurs à une autre. Cette perspective relègue la notion de transparence au statut de principe éthique flottant dépourvu de toute utilité pratique. Ce flou artistique contribue à rendre peu contraignant cet impératif et à limiter les interventions de design à quelques aménagements explicatifs dont l'efficacité doit encore être prouvée.

Plus important encore, cet impératif de transparence, tel qu'il est actuellement formulé, maintient la recherche dans une dynamique réactive plutôt que transformative. Nous pouvons douter de l'efficacité d'une démarche qui vise à corriger à posteriori un dispositif initialement construit selon une logique de disciplinarisation¹⁵. Rendre transparent n'implique pas de donner accès à une réalité objective mais est avant tout une opération de médiation qui occulte autant qu'elle n'éclaire. Cette démarche fait partie intégrante de la conception de l'interface du dispositif et, par conséquent, de la définition de sa fonction. Elle doit dès lors être définie pour un public spécifique et viser un objectif précis de sorte qu'il soit possible de formaliser clairement ce qui doit être mis en lumière. La transparence est donc un moyen pour une finalité qui doit être définie en fonction d'une problématique. Aussi maladroite soit-elle, la notion de transparence a toutefois pour mérite d'avoir fédéré, au sein de la littérature scientifique, une série de méthodes qui sont autant de pièces qu'il

¹⁵ A ce sujet, nous renvoyons au premier chapitre de cette thèse.

est possible d'articuler entre elles pour composer l'interface d'un système technique : visualisations de données, outils de contrôles, modalités d'explications etc. Nous reviendrons sur ces éléments dans le chapitre 3. Il reste donc à circonscrire une finalité précise qui dépasse la simple mise en lumière du fonctionnement de l'algorithme. Afin de décrire ce que pourrait être cette finalité, nous proposons de travailler à partir du concept d'*autonomie*. Le choix de cette notion est surtout motivé par sa récurrence dans la littérature que nous venons de parcourir. Intuitivement, l'autonomie renvoie à la capacité de l'utilisateur à garder un contrôle plus ou moins étroit de l'algorithme et des contenus qu'il consomme. Toutefois, cette perspective nous semble trop étroite et invalide de nombreux usages qui reposent sur le maintien d'un certain degré d'incertitude. En analysant les pratiques d'écoute d'une plateforme de streaming musical, Beuscart *et al.* (2019) observent ainsi que :

Cette délégation est proche de celle du choix d'une station de radio dans le cas du flux, plus ponctuelle dans le cas d'un suivi de lien de recommandation. Ces outils semblent tenir leurs promesses d'exploration de la richesse des catalogues et d'enrichissement de la diversité consommée des utilisateurs, en les guidant préférentiellement vers des titres de longue traîne ou de notoriété intermédiaire. Il reste qu'il est difficile de qualifier systématiquement d'hétéronomes et de passives ces écoutes où l'agentivité algorithmique est plus marquée. Le suivi de lien de recommandation (« vous avez aimé tel artiste, vous devriez aimer tel autre ») peut aussi bien être vu comme une pratique de novice (comblant son incompetence par la délégation à l'algorithme) que comme une démarche d'amateur éclairé (utilisant ponctuellement un outil informatique pour compléter sa palette exploratoire). (p.43)

Au sein de la littérature scientifique, on la retrouve en interaction homme-machine (B. Friedman, 1996), dans le domaine des systèmes de recommandation (Beuscart *et al.*, 2019 ; Knijnenburg *et al.*, 2016), dans la littérature consacrée aux algorithmes de service public (Bernstein *et al.*, 2021 ; Milano *et al.*, 2020) et en éthique appliquée de l'intelligence artificielle (Floridi *et al.*, 2018). Elle semble donc mobiliser un intérêt transversal à plusieurs domaines. Toutefois, malgré cette récurrence, la notion reste assez vague et traduit des préoccupations qui ne se croisent que partiellement. Elle est parfois définie comme un problème de charge cognitive, d'actualisation d'un potentiel ou encore de délégation d'autorité. L'établissement d'un vocabulaire plus formel nous

semble nécessaire pour pouvoir articuler ces différentes perspectives et former un modèle capable de décrire avec plus de finesse la nature distribuée de notre agentivité. Ce n'est qu'une fois que nous aurons fixé un cadre formel autour de cette notion, nous permettant ainsi d'opérer une distinction claire entre usage autonome et hétéronome, que nous pourrons entamer notre réflexion sur les modalités d'interaction homme-machine plus à même de soutenir cette autonomie.

Dès lors, dans la suite de ce chapitre, nous nous attacherons, dans un premier temps, à désambigüiser ces différentes conceptions de l'autonomie tout en les resituant dans un paysage conceptuel plus large. Dans un second temps, à travers la mise en perspective critique de ces différentes notions, nous formulerons une conception de l'autonomie qui puisse être opérationnalisée dans le cadre du design d'un algorithme de recommandation de service public.

AUX ORIGINES DE L'AUTONOMIE¹⁶

Preamble

L'usage courant de l'autonomie renvoie bien souvent à une faculté d'action intentionnelle, qu'elle soit motrice ou intellectuelle. La gestion de l'autonomie dans le domaine de la santé renverra donc au soutien de l'individu dans la réalisation de l'une ou l'autre tâche qu'il n'est pas ou plus capable de réaliser. Appliqué à l'enfance, l'emploi du terme indique également une incapacité à réaliser certaines tâches et à vivre par ses propres moyens. La figure de l'adulte en bonne santé est l'idéal-type de cette vision de l'autonomie. Dans un contexte technique, on parlera de système autonome pour désigner un appareillage technique capable de réaliser une série d'actions sans la supervision directe d'un être humain. On parle ainsi de voiture autonome pour désigner un véhicule capable de se diriger sans être préalablement contrôlé par un conducteur humain. Nous ne faisons pas ici référence à ces conceptions communes de l'autonomie. Enfin, la notion d'autonomie est souvent mobilisée dans un contexte politique pour désigner la capacité d'une institution ou d'une collectivité à gérer ses affaires internes sans interférence externe. En ce qui nous concerne, nous faisons plutôt référence à une autre conception assez répandue qui décrit la capacité d'un individu à donner une certaine forme ou une direction précise à la vie qu'il mène. Il façonne d'une façon ou

¹⁶ L'historique que nous réalisons de l'autonomie doit beaucoup aux travaux de Hinchman (1996) et Mackenzie et Stoljar (2000) dont nous reprenons les grandes catégories historiques, respectivement pour la période s'étendant entre Kant et les perspectives procéduralistes et pour les travaux féministes plus récents.

d'une autre ses actions selon une série de principes ou de contraintes qu'il s'impose plus ou moins volontairement.

Etant donné la multitude de contextes dans lequel ce concept peut être mobilisé, il est particulièrement difficile d'en faire un état de l'art complet. A cette première difficulté vient s'ajouter le nombre important de notions connexes (telles que l'agentivité, la raison, le libre-arbitre, l'esprit critique etc.) qui viennent se fondre et se confondre avec ce concept. Il en résulte une démultiplication des sens associés à cette notion selon le problème et la perspective adoptée. Dworkin (2017) nous donne une idée générale de cette entreprise qui retrace plusieurs siècles de philosophie morale :

There is a Kantian notion of autonomy as the supreme principle of morality and a possible Kantian notion of autonomy as freedom. There is autonomy as the use of practical reason in the pursuit of happiness. There is autonomy as an ideal that the state ought to foster. There is Rawlsian autonomy (In the Theory of Justice), which is the ability of a person to limit his pursuit of the good to the right. There is Rawlsian rational autonomy (in Political Liberalism), which is a power to revise and pursue a conception of the good. There is Razian autonomy, which is an attempt to engage with the good which seems to have as an implication that the exercise of personal morality already involves the exercise of a particular moral autonomy. There is personal autonomy as being responsive to one's conscience. And finally, there is the possibility of an intermediate position between personal and moral autonomy. (p.450)

Plutôt que de tenter l'exercice périlleux de fournir un état des lieux général des travaux portant sur cette notion, il nous semble beaucoup plus pertinent, comme le suggère Dworkin (2017), de nous interroger sur certaines dimensions spécifiques du concept en regard d'un problème clairement identifié. En l'occurrence, comment penser l'autonomie d'un individu dans sa relation à des outils prédictifs automatisés tels que des algorithmes de recommandation ? Si une certaine sérendipité a bien guidé le processus de sélection des concepts présentés, deux impératifs ont toutefois structuré nos recherches :

- 1) Un impératif pragmatique : Notre premier souci fut de construire un assemblage théorique qui nous permette de garder en vue les pratiques réelles des utilisateurs mais également d'expliquer et de contextualiser des pratiques de design existantes. Nous avons dès lors mis

de côté certains penseurs de philosophie morale dont les travaux s'orientent vers des problématiques générales de justice et de moralité pour privilégier les travaux portant sur l'autonomie personnelle.

- 2) Une cohérence interne : Comme le témoigne notre recours aux travaux de Foucault, Verbeek ou encore Hutchins, notre travail repose sur une ontologie relationnelle au sein de laquelle les relations entre objets et sujets sont à la fois complexes et multidirectionnelles. Dans le cadre de la critique de cette revue de littérature, nous avons dès lors opté pour des théories avec un parti pris ontologique similaire.

Dans les pages qui suivent, nous commencerons par décrire la notion d'autonomie en passant par la philosophie grecque, la pensée de Kant et la psychologie humaniste de sorte à dégager les principaux courants qui irriguent encore aujourd'hui les conceptions les plus communes de ce concept. La suite de notre analyse visera à identifier les limites inhérentes à ces approches. C'est notamment à travers l'exploration de la notion de souci de soi, telle qu'elle est présentée par Foucault, que nous tenterons d'identifier une autre trajectoire d'analyse qui nous permette de réhabiliter une conception de l'autonomie qui soit attentive à la nature socialement et matériellement distribuée de notre agentivité.

Origines grecques

La notion d'autonomie est ancrée dans l'antiquité grecque. Son étymologie provient du grec ancien *autos* pour « soi » et *nomos* pour « règle ». Ces origines grecques peuvent être retrouvées dans les textes de Platon et d'Aristote. Une première origine peut être identifiée dans la notion aristotélicienne d'*autarkeia* ou d'auto-suffisance. Le terme est employé pour désigner des cités états auto-suffisantes capables de se gouverner par elles-mêmes et de se préserver de pouvoirs extérieurs. Cette idée d'auto-suffisance se retrouve également dans la conception aristotélicienne de grandeur d'âme. Un individu répondant à ce qualificatif est capable de poursuivre des objectifs qui ne découlent pas de la pression exercée par le besoin ou l'utilité (May, 1994).

Une autre origine peut être identifiée dans la maîtrise de soi platonicienne. L'homme est conceptualisé comme un enchevêtrement complexe de dimensions supérieures et inférieures. La maîtrise de soi désigne le processus par lequel les dimensions rationnelles de l'homme contribuent à l'élaboration de relations de sujétion adéquates entre ces éléments supérieurs et inférieurs. Comme le fait remarquer Gibbs (1979), ces relations de sujétion ne doivent pas être comprises

comme une forme d'auto-répression mais plutôt la création d'une harmonie interne propre à ce que Platon appelle l'homme tempérant. La maîtrise de soi implique l'exercice d'une sagesse, située dans la dimension rationnelle de l'homme, qui corrigera pulsions, désirs et autres inclinations de sorte à ce qu'elles tendent vers ce qui est bon pour l'individu. L'homme intempérant, à l'inverse, est dirigé par ses passions et non par une raison fonctionnelle (Gibbs, 1979). Cette conception de l'autonomie apparaît par la suite chez les stoïciens. Devenir auto-suffisant implique de se détacher de ses passions de sorte à ce qu'elles ne puissent pas diriger l'action (Hinchman, 1996).

L'autonomie morale de Kant

Les travaux de Kant sur la morale sont souvent considérés comme une étape fondamentale dans l'évolution de la réflexion occidentale sur l'autonomie (Hinchman, 1996). Kant emploie ici le terme d'autonomie comme le respect à l'égard d'une règle morale que s'impose soi-même un individu. Cette contrainte morale peut être considérée comme objective car elle découle de l'impératif catégorique développé par Kant : « Agis de telle sorte que le principe de ton action puisse être érigé en loi universelle » (Kant, 1788/2003). Une règle est universalisable si n'importe quel individu doué de raison est amené à formuler la même règle à travers son propre exercice de la raison. L'autonomie kantienne est donc intimement liée à la problématique de la morale. Les principes moraux ne découlent pas de la nature mais de la raison. Toute forme de choix qui repose sur des émotions ou d'autres considérations quotidiennes est considérée comme hétéronome.

Il n'est pas possible pour nous de rendre justice à cette perspective et d'en extraire la notion d'autonomie sans dénaturer la métaphysique morale que Kant lui associe. De plus, en dehors de la philosophie politique, les travaux contemporains sur l'autonomie personnelle ne trouvent presque plus d'écho dans cette définition formelle de l'autonomie. Toutefois, il est possible d'identifier une autre conception de la morale chez Kant à travers son essai « Qu'est-ce que les lumières ? » qui partage beaucoup plus de similitudes avec les conceptions modernes de l'autonomie et qui semble avoir laissé une empreinte plus importante dans l'imaginaire collectif en dehors de la sphère académique (Hinchman, 1996). En 1784, le périodique allemand « Berlinische Monatschrift », après avoir posé à son lectorat la question « Was ist Aufklärung?¹⁷ », publie la réponse de Kant. Le texte est particulièrement mineur par rapport à l'ensemble de l'œuvre de Kant mais, comme le

¹⁷ Qu'est ce que les Lumières ?

constate Foucault (2015), il y développe une réflexion ancrée dans l'actualité politique de son époque. Dans cet essai, Kant définit Les Lumières comme un moment de sortie de l'homme d'un état d'immaturité (Kant, 1784/2020). Cet état, qu'il nomme minorité, désigne un état d'assujettissement de notre volonté à l'autorité d'un précepteur dans des domaines où notre propre raison devrait s'exercer. Cette minorité découle de l'incapacité de l'homme à faire preuve de son propre entendement dans la conduite de sa vie. Ceux qui, par contre, sont capables de prendre la responsabilité de leurs actes sans recourir à la guidance d'autrui atteignent un état de majorité. Cet appel à la majorité est toutefois nuancé par la distinction qu'opère Kant entre un usage public et un usage privé de la raison. L'exercice de cette liberté de raisonnement ne peut se faire que dans le cadre d'un usage public de la raison, c'est-à-dire en dehors de toute fonction publique lorsque l'on s'adresse à l'oral ou à l'écrit à un public. L'usage privé réfère à l'usage qui est fait lors de l'exercice d'une fonction civile (un officier en service ou un prêtre lorsqu'il s'adresse à sa paroisse). Dans ces cas-là, il faut obéir afin de maintenir l'ordre. Si l'usage public de la raison doit être absolument libre pour répandre l'idéal des Lumières, l'usage privé lui, peut être étroitement limité. Ainsi, un officier n'est pas en droit de contester les ordres qu'il reçoit au moment où il doit les appliquer. Toutefois, rien ne l'empêche de formuler une critique des erreurs propres à son service qu'il soumettrait ensuite à un public.

Même si Kant n'emploie pas ici le terme d'autonomie, ce qu'il décrit comme un état de majorité partage de nombreuses similitudes avec une conception moderne de l'autonomie entendue à la fois comme absence de contraintes extérieures mais aussi comme capacité d'être l'auteur de ses propres lois. Toutefois, le lexique kantien est avant tout consacré à la conceptualisation d'une autonomie morale et laisse assez peu de place à la formalisation d'une autonomie personnelle focalisée sur les choix du quotidien (Diana T. Meyers, 1989). L'autonomie kantienne est particulièrement présente dans certaines représentations de l'éducation. Dans son « Vocabulaire technique et critique de la pédagogie et des sciences de l'éducation » publié en 1974, Leif (1974) décrit l'autonomie comme :

[...] la capacité de l'être humain lui permettant de se déterminer selon les normes de sa propre réflexion et de sa propre volonté. Elle se manifeste par l'indépendance de la conduite, celle-ci étant régie conformément à des règles ou des impératifs que le sujet se donne à lui-même, par son libre choix. (p.31)

Bien que cette perspective de l'autonomie qui valorise la maîtrise de la raison est assez connue, d'autres approches seront plus ouvertes aux signaux émotionnels et seront centrées sur l'identification des « besoins profonds » de l'individu.

Romantisme et psychologie humaniste

Hinchman (1996) identifie, dans le courant romantique du 19ème siècle, une autre origine importante de notre conception contemporaine de l'autonomie. Plutôt que de proposer un protocole d'universalisation de sa posture morale, l'approche romantique conceptualise l'autonomie comme un mouvement vers soi qui doit nourrir une recherche d'individualité. La raison n'est plus le chef d'orchestre qui vient mettre au pas un ensemble désordonné d'émotions et de pulsions primaires irrationnelles. Ces pulsions, qui doivent être canalisées plutôt que maîtrisées, deviennent la base fondamentale d'une individualité.

Cette approche se retrouve notamment dans la démarche de John Stuart Mill et plus spécifiquement dans son livre « On Liberty » :

« Strong impulses are but another name for energy. Energy may be turned to bad uses; but more good may always be made of an energetic nature, than of an indolent and impassive one. Those who have most natural feeling are always those whose cultivated feelings may be made the strongest. The same strong susceptibilities which make the personal impulses vivid and powerful, are also the source from whence are generated the most passionate love of virtue, and the sternest self-control. It is through the cultivation of these that society both does its duty and protects its interests: not by rejecting the stuff of which heroes are made, because it knows not how to make them. A person whose desires and impulses are his own – are the expression of his own nature, as it has been developed and modified by his own culture – is said to have a character. One whose desires and impulses are not his own has no character, no more than a steam engine has a character ». (Mill, 1956, p. 56)

Contrairement à l'approche kantienne qui repose sur l'usage libre de sa raison pour universaliser ses maximes, l'approche romantique repose plutôt sur le développement de son soi personnel. Comme le fait remarquer (Hinchman, 1996):

The idea is precisely not to adopt a universalizing stance toward one's own moral and emotional life, to « get outside » the self toward some more impartial, truer vantage point, but instead to externalize what has always lain within, albeit only dimly recognized, as the unique, noninterchangeable core of self. (p.498)

Être autonome n'est donc plus la clé de voûte d'une action morale mais plutôt la pratique d'expression de soi nécessaire à la constitution de son identité.

L'association du développement de soi à la problématique de l'autonomie peut également être retrouvée dans la psychologie humaniste américaine et notamment dans les travaux d'Abraham Maslow et Carl Rogers. Selon cette perspective, la finalité du développement personnel n'est autre que l'acte de devenir une personne à travers un processus de *réalisation de soi* (*self-actualization*). Ces théories sont caractérisées par une sorte « d'optimisme moral » quant à la nature humaine et par une dimension normative qui situe le devenir de l'homme dans une dynamique de développement personnel individuel (Geller, 1982). En bref, elles décrivent une feuille de route du développement humain idéal. La nature de l'individu est bonne par essence et la finalité de son existence réside dans la réalisation de ce potentiel, qui lui apportera épanouissement et contentement (Geller, 1982). Selon Rogers (1966/1995), devenir un individu authentique implique un mouvement vers soi d'identification de sa nature profonde. S'accomplir implique de vivre en harmonie avec qui nous sommes réellement. Ce soi est déjà construit et il s'agit donc de trouver le chemin qui y mène à travers un travail d'introspection.

La théorie de la motivation de Maslow (1954), pour sa part, repose sur l'identification et la distinction entre les besoins primaires, comme la faim, et les besoins de plus haut niveau, comme l'épanouissement personnel. La satisfaction des besoins de plus bas niveau est nécessaire à la réalisation des besoins de plus haut niveau qui correspondent à la réalisation pleine et totale du potentiel humain. Ce n'est qu'une fois libéré des contraintes imposées par ces besoins d'ordre inférieur que l'individu peut réellement devenir autonome et s'investir dans des entreprises d'un ordre, supposément, supérieur comme les arts ou la science (Maslow, 1954).

Nous proposons ici une petite parenthèse dans notre tracé historique pour revenir au domaine des systèmes de recommandation. Les théories psychologiques humanistes ont connu un regain d'intérêt dans le domaine depuis quelques années avec le réemploi de la notion de réalisation de

soi issue des travaux de Maslow. L'expression apparaît pour la première fois dans un article théorique de Knijnenburg *et al.* (2016):

By making us better understand our own preferences, Recommender Systems for Self-Actualization will improve our potential to have confidence in (and take ownership over) our life decisions. They allow us to each develop a unique personal style, thereby supporting Maslow's need for Esteem and Self-Actualization, and preventing the erosion of our autonomy as consumers. (p.13)

Les auteurs mobilisent explicitement les travaux de Maslow en réponse à la problématique de la bulle de filtres de Pariser (2011). L'apport de la psychologie humaniste est particulièrement peu contextualisé et la mention des travaux de Maslow n'apparaît que dans la conclusion de l'article. A la suite de cet article, plusieurs autres travaux reprennent cette notion de réalisation de soi en citant Knijnenburg *et al.* (2016) sans pour autant rentrer dans plus de détails théoriques (Graus *et al.*, 2018 ; Harambam *et al.*, 2019 ; Sullivan *et al.*, 2019 ; Wilkinson, 2018). Le recours à la perspective humaniste n'est pas surprenante lorsque l'on constate que les deux démarches, à la fois celle des psychologues humanistes des années 40 et celle de Knijnenburg *et al.* (2016) se construisent en opposition à une démarche behavioriste.

La théorie de Maslow repose notamment sur une conception du sujet et de ses besoins comme une structure innée. Si personne ne remet en doute qu'un besoin primaire tel que la faim est bien innée, on peut toutefois s'interroger sur le statut de besoins plus complexes. Selon Neher (1991), Maslow cherche visiblement à déprécier l'influence de facteurs environnementaux sur le développement de la psychologie humaine en réponse à l'erreur inverse attribuée aux behavioristes qui consiste à survaloriser cette influence. Pour reprendre les propos de Maslow :

Once [lower needs are met] each person proceeds to develop in his own style, uniquely, using these necessities for his own private purposes. In a very meaningful sense, development then becomes determined from within rather than from without... The role of the environment is ultimately to permit him or help him to actualize his own potentialities [because] he knows better than anyone else what is good for him. (Maslow cité dans Neher, 1991 p.93)

La même année que Knijnenburg *et al.* (2016), Ekstrand et Willemsen (2016) publient un article intitulé « Behaviorism is not enough ». Bien que les auteurs ne fassent pas directement référence à la psychologie humaniste, ils rédigent dans cet article une critique de l'usage du behaviorisme dans le domaine de la recommandation automatisée¹⁸. A travers l'usage presque exclusif de signaux implicites, le système repose sur une interprétation partielle du comportement de l'utilisateur sans lui laisser l'opportunité de spécifier des objectifs personnels. Le dispositif ne peut donc faire la différence entre une action motivée par un intérêt occasionnel d'une action motivée par une addiction (Ekstrand et Willemsen, 2016). Un design plus vertueux doit, selon les auteurs de l'article, aider l'utilisateur à devenir une meilleure version de lui-même.

Dans l'ensemble, ces travaux reposent principalement sur un principe de connaissance de soi en opposition à un design de l'engagement et de la gratification instantanée. Les propositions de design qui en découlent visent à faire du système de recommandation un outil de découverte de son « soi intérieur » à travers 3 leviers principaux :

- 1) La mise en évidence des éléments saillants du profil de l'utilisateur et l'explication de leur effet sur les recommandations produites (Graus *et al.*, 2018) ;
- 2) L'ajout d'outils de contrôle de l'algorithme (Harambam *et al.*, 2018) ou la mise à disposition de principes alternatifs de recommandations du type « recommandations que vous allez peut être détester », « recommandations sur lesquelles nous ne sommes pas sûr », « recommandations que vous serez le premier à expérimenter » etc. (Knijnenburg *et al.*, 2016 ; Knijnenburg *et al.*, 2016) ;
- 3) La permission accordée à l'utilisateur d'exprimer ses besoins de façon explicite (Ekstrand et Willemsen, 2016).

Les travaux de Maslow et de Rogers peuvent être critiqués sur plusieurs points mais ce n'est pas l'objet de cette recherche. Toutefois, il nous semble pertinent de souligner la critique la plus récurrente de ces théories concernant leur conception problématique du sujet. Rogers postule un sujet intérieur, une substance immatérielle, inaltérable et dont la genèse est antérieure à toute forme

¹⁸ Pour une explication plus détaillée de l'influence de ce paradigme dans le design de systèmes de recommandation, nous renvoyons le lecteur au premier chapitre de cette thèse et, plus précisément, à la section consacrée à l'évolution historique de ces technologies.

de conscientisation (Geller, 1982). De son côté, Maslow postule un soi implicite qui découle d'un noyau génétique et qui se déplie en fonction de conditions environnementales plus ou moins favorables à l'épanouissement personnel (Geller, 1982). Autant chez Maslow que chez Rogers, le sujet « véritable » est un sujet atomisé et asocial. Cette perspective rentre évidemment en contradiction avec la position développée dans le premier chapitre qui postule un sujet construit à travers le jeu de ses relations sociales et techniques. (Geller, 1982 ; Neher, 1991)

Dans les faits, les propositions techniques issues des sciences informatiques s'écartent quelque peu de l'inspiration théorique. Cela s'explique probablement par l'héritage de plus de dix ans de travaux en expérience utilisateur qui restent plus influents qu'une théorie psychologique introduite plus tardivement. Ainsi, la proposition de Knijnenburg *et al.* (2016) de fonder la recommandation comme support plutôt que comme substitut à la prise de décision mérite d'être retenue pour la suite de notre entreprise :

Traditional recommenders turn preferences into choice options, but research shows that user preferences are fleeting, constructed on the fly and vulnerable to distorting influences, rather than welldefined, fixed, and invariant. RSSAs¹⁹ take the additional step to help users develop and express their preferences. (Knijnenburg *et al.*, 2016, p. 12)

Il est également possible de discerner un semblant de stoïcisme, notamment chez (Ekstrand et Willemsen, 2016). Là où les abus du behaviorisme conduit à l'exploitation des biais de l'individu, un système plus éthique doit lui permettre de s'en détacher. Autonomie procédurale

Les approches les plus récentes de l'autonomie au sein de la philosophie morale sont marquées par un retour à l'authenticité romantique et à son mélange avec le principe kantien d'exercice personnel de son entendement (Hinchman, 1996). On y retrouve à la fois l'idée d'écouter une « voix intérieure » tout en appliquant rigoureusement un raisonnement réflexif et critique. Les formulations contemporaines de l'autonomie sont principalement des variantes d'une perspective dite procédurale. En réaction à une conception substantive de l'autonomie dotée d'un caractère normatif prononcé qui promeut l'une ou l'autre conception morale, certains philosophes se sont attachés à formaliser l'autonomie sous la forme d'une procédure cognitive indépendante du contenu sur lequel porte la réflexion et la prise de décision. L'important n'est donc plus la nature

¹⁹ Recommender System for Self-Actualization

morale, ou non, de la règle que l'individu s'impose mais plutôt le processus de réflexion critique qui a conduit à l'élaboration et l'application de cette loi. Ces approches cherchent donc à identifier les conditions nécessaires et suffisantes pour qu'une décision, un engagement, une valeur puisse être considérée comme autonome à la suite de ce processus de réflexion critique. Réaliser cette procédure en satisfaisant les conditions énoncées doit permettre de garantir que le choix individuel est authentique, c'est à dire qu'il correspond aux volontés d'un « vrai soi ».

Au sein des théories procédurales, la plus connue est probablement la théorie hiérarchique telle qu'elle a pu être formulée par Frankfurt (1971) ou Dworkin (1976). Les théories hiérarchiques reposent sur l'identification et la structuration d'une série d'éléments composant le soi. Frankfurt et Dworkin postulent un individu divisé entre un soi de bas niveau et un « vrai » soi qui dirige le processus de réflexion d'un « soi total » (M. A. Friedman, 1986). Selon ces perspectives, l'autonomie d'un individu ne peut s'exprimer que par sa capacité à formuler des jugements et des préférences sur des désirs. Frankfurt (1971) définit l'autonomie comme un accord entre les désirs de premier ordre d'un agent et ses volitions de second ordre. Les désirs de premier ordre composent ce que l'on associe traditionnellement à la volonté d'un individu et ce qui le motive à agir (avoir l'envie de dormir, de manger, de voyager etc.). Un désir de second ordre est un désir portant sur un désir. Je peux, par exemple, désirer être un amateur de course à pied, car cela m'encouragerait à faire du sport plus souvent. Le désir ne porte pas sur la course à pied mais sur le souhait de désirer apprécier l'activité. Un désir de second ordre devient une volition de second ordre lorsque l'on souhaite qu'un désir devienne effectif et qu'il dirige sa volonté. Toujours selon Frankfurt (1971), être une personne implique d'être capable de poser un jugement sur la « désirabilité » de ses désirs de sorte à définir ceux que l'on souhaite rendre effectifs. Être autonome implique d'assurer la conformité entre ses volitions de second degré et ses désirs de premier degré. Pour expliquer ceci, Frankfurt (1971) prend l'exemple du toxicomane volontaire et involontaire. Le toxicomane volontaire est considéré comme autonome car, même s'il n'était pas soumis à son addiction, il souhaiterait tout de même vivre une vie dans laquelle il consomme et apprécie consommer certaines drogues. A l'inverse, le toxicomane involontaire n'est pas autonome car ses désirs de second degré sont en discordance avec ses désirs. Il souhaite être motivé par des désirs différents mais il n'y arrive pas. Selon cette perspective, la nature de l'objet sur lequel porte un désir est sans importance pour désigner le caractère autonome ou non d'un individu. Une motivation devient réellement sienne lorsque l'individu juge qu'il souhaite bel et bien être mis en action par ces motivations,

lorsqu'il le désire « de tout cœur ». Ces perspectives accordent dès lors un statut particulier à la réflexion critique qui serait donc plus représentative de « soi-même », contrairement à des motivations de premier ordre non évaluées (M. A. Friedman, 1986).

Authenticité

Ces approches définissent un comportement autonome comme une prise de décision authentique. Une décision est authentique quand, à la suite d'un processus réflexif interne à l'individu, elle correspond à ses désirs sincères. Chacune de ces approches diverge en fonction de la structure du processus d'évaluation ainsi que de la nature de ce « soi » authentique et réflexif.

Selon une perspective procédurale à la Dworkin/Frankfurt, l'authenticité de soi est garantie par la cohérence interne de la volonté de l'individu. Les théories hiérarchiques furent notamment critiquées pour leur rigidité au sein de la structure interne de l'individu plaçant la barre particulièrement haut pour quiconque souhaiterait être qualifié d'autonome (Mackenzie, 2019). Comme le souligne très justement Mackenzie (2019), alors que Frankfurt associe l'ambivalence à une « maladie de l'esprit » (Frankfurt, 1999, p. 100), un certain degré d'incertitude et de conflit psychologique est à la fois inévitable mais parfois également nécessaire. Diana T. Meyers (1989) critique également ces approches pour leur rigidité qui impose au sujet d'en connaître plus sur lui-même qu'il ne pourrait probablement jamais réellement savoir. De plus, si l'influence de la socialisation de l'individu sur la formation de ses valeurs et engagements est bien reconnue, le processus de réflexion critique doit permettre à la raison de transcender cette nature sociale. Dès lors, ces théories hiérarchiques ne permettent pas d'expliquer en quoi une motivation de second ordre est ontologiquement plus proche d'un soi authentique qu'un désir de premier ordre (Mackenzie et Stoljar, 2000). Ces critiques ont amené plusieurs théoriciens à développer une perspective qui se veut plus complète à l'égard du processus de formation d'une identité.

Friedman prend notamment l'exemple d'une femme au foyer des années 50 en pleine contradiction (M. A. Friedman, 1986). D'une part, cette femme au foyer a complètement intégré la norme sociale selon laquelle une mère de famille convenable doit rester au foyer et placer ses propres besoins au second rang de ses préoccupations pour privilégier les besoins de sa famille. D'autre part, elle ne peut s'empêcher de ressentir des épisodes de frustration lors desquels elle ressent certains besoins en contradiction avec les normes sociales dont elle reconnaît la totale validité. Étant donné l'intégration de valeurs et de principes reposant sur le sacrifice de soi en tant qu'idéal féminin, elle

ne dispose pas de principes « d'un ordre supérieur » à ses motivations de premier ordre qui lui permettent de formuler une critique de sa condition. Pour Friedman, ces épisodes émotionnels caractérisés par des sentiments de frustration, de deuil ou de dépression sont de biens meilleurs indicateurs des besoins authentiques de l'individu qu'un processus de réflexion rationnel qui est également soumis aux effets de sa socialisation. C'est pour cette raison que Friedman propose un modèle d'intégration qui vise à s'éloigner d'une structuration hiérarchique plaçant la raison au sommet du processus réflexif en remplaçant le modèle top-down de Dworkin/Frankfurt par une dynamique à double sens. En même temps que des principes d'un degré supérieur permettent d'évaluer ses motivations de premier degré, ces mêmes motivations de degré inférieur sont un autre point de vue depuis lequel il est possible d'examiner ces principes d'ordre supérieur.

Friedman met en parallèle son approche avec les processus de raisonnement inductif et déductif. Si l'approche hiérarchique correspond à un processus déductif basé sur des prémisses théoriques, son modèle à double sens vise à compléter cette approche par une dynamique inductive qui opère depuis le « terrain » :

Scientific explanation involves the process by which covering laws and explanatory principles are inducted from what they are supposed to cover, namely, particular phenomena. My analogous proposal is that autonomy involves the process by which a person's « covering » principles, especially those at the highest level, are « inducted » from what they are supposed to cover, namely, lower level standards and motivations. Revising one's principles in the light of one's motivations constitutes an « inductive » leap. And autonomy necessitates such leaps as much as it does the use of those new principles of one's own to inform the critical assessment of (future) motivation. (M. A. Friedman, 1986, p. 33)

On peut toutefois reprocher à ces approches, autant celle de Friedman que de Dworkin et Frankfurt, de limiter le processus réflexif à l'instant de la prise de décision et d'ignorer par conséquent les conditions historiques de production du sujet qui contribuent à déterminer ses préférences, ses valeurs ou encore ses engagements personnels. Or, la connaissance de l'historicité de ses engagements peut amener un individu à réviser sa position et à rejeter des engagements qui lui apparaissent maintenant inauthentiques. Dworkin propose dès lors une perspective historique, malgré tout très similaire à l'approche de Frankfurt reposant sur une condition d'indépendance

procédurale. Un individu est autonome si, à travers un travail de réflexion critique, il finit par approuver les différents processus, comme l'influence de pairs ou la pression parentale, qui l'ont amené à formuler certaines valeurs ou engagements (Mackenzie et Stoljar, 2000). Autrement dit:

... in most cases we cannot be said to have adopted or chosen or selected our beliefs, desires, emotions, principles, and so forth. In some cases this idea doesn't even make sense. In other cases it is simply a contingent fact that we find ourselves moved in certain ways. Autonomy cannot be located on the level of first-order considerations, but in the second-order judgments we make concerning first-order considerations. If the autonomous man cannot adopt his motivations de novo, he can still judge them after the fact. The autonomous individual is able to step back and formulate an attitude towards the factors that influence his behavior. (Dworkin, 1976, p. 24)

Christman (2009), pour sa part, est bien conscient des limites inhérentes à une telle démarche d'introspection. Il est généralement très rare qu'un individu soit complètement honnête avec lui-même de façon délibérée, constante et consciente (Mackenzie, 2019). Pour pallier cela, Christman propose une condition d'authenticité plus accessible que le processus d'identification proposé par Dworkin et Frankfurt. Les engagements pris par l'agent doivent passer un test de non aliénation :

To be alienated is to experience negative affect, to feel repudiation and resistance. It is, moreover, to feel a need to repudiate that desire or trait, to reject it and alter it as much as possible, and to resist its effects. If, while feeling alienation and self-repudiation of this sort, a person is unable to rid herself of the characteristic in question, she is heteronomous in relation to it. (Christman, 2009, p. 144).

Cette condition de non-aliénation déborde du domaine purement cognitif pour inclure une dimension somatique. Au-delà d'un travail rationnel d'introspection et de délibération, ce processus d'auto-évaluation critique implique d'être à l'écoute de son corps et de ses réactions émotionnelles. Contrairement à l'identification qui demande une acceptation des traits évalués, le principe de non-aliénation demande seulement que l'individu ne ressente pas un sentiment de rejet à leur égard après avoir identifié leurs conditions historiques et sociales. Il peut donc y être indifférent ou ne pas avoir encore d'avis clair sur la question.

Ce processus de réflexion critique et de révision de ses engagements ne doit pas, selon Christman, s'exercer tout le temps, ce qui serait tout aussi oppressif. Si certaines dimensions de notre être sont plus à même de subir des révisions et des altérations fréquentes, certains attachements et ancrages émotionnels définissent de façon beaucoup plus profonde l'identité d'une personne. Christman prend comme exemple l'amour d'un parent pour son enfant comme exemple d'attachement qui est, a priori, beaucoup plus difficile à renégocier. L'exercice d'un contrôle permanent sur de telles dimensions est virtuellement impossible et empêcherait tout projet de plan de vie. Dès lors, Christman considère comme irréaliste toute notion de l'autonomie qui demande une telle capacité de révision. C'est notamment l'existence d'un conflit entre le registre rationnel et émotionnel qui doit signaler à l'individu l'existence d'une non authenticité dans ses actions (Christman, 2004).

Les approches de l'autonomie que nous avons présentées jusqu'à maintenant, principalement l'approche issue de la psychologie humaniste et l'approche procédurale de Frankfurt, reposent sur une recherche d'authenticité, d'un soi profond avec lequel il faudrait renouer à travers un exercice ponctuel d'introspection. Toutefois, l'ensemble du cadrage théorique mobilisé dans le premier chapitre, qui nous sert à contextualiser la nature de l'agentivité humaine, repose sur un sujet fragmenté et distribué. Autant la théorie de la cognition distribuée qu'une approche foucaldienne du pouvoir, les deux théories principales sur lesquelles repose notre analyse du pouvoir algorithmique, dessinent un sujet dont l'identité et l'agentivité sont hautement conditionnées par l'environnement matériel et social. En fonction des approches, s'extraire de ces relations d'interdépendance est soit impossible, soit contre-productif. Dès lors, les approches présentées jusqu'à maintenant, notamment les travaux issus de la psychologie humaniste, peuvent difficilement être mobilisées dans le cadre de nos travaux. Il est dès lors nécessaire de trouver une voie intermédiaire qui nous permette de penser cette notion d'autonomie dans un contexte d'agentivité distribuée.

SOUCI DE SOI

Malgré que Foucault n'emploie pas la notion d'autonomie, ses travaux sur le sujet nous semblent toutefois intéressants à explorer pour envisager une autre voie de l'autonomie qui ne repose pas sur l'introspection comme finalité mais plutôt sur une pratique à la fois éthique et esthétique. Il éprouve un certain scepticisme à l'égard de cette approche de l'authenticité comme idéal du rapport à soi :

Le thème de l'authenticité renvoie explicitement ou non à un mode d'être du sujet défini par son adéquation à lui-même. Or il me semble que le rapport à soi doit pouvoir être décrit selon les multiplicités de formes dont l'« authenticité » n'est qu'une des modalités possibles ; il faut concevoir que le rapport à soi est structuré comme une pratique qui peut avoir ses modèles, ses conformités, ses variantes, mais aussi ses créations. La pratique de soi est un domaine complexe et multiple. (Foucault, 1999a, p. 617)

A d'autres occasions, il oppose de façon plus sévère ce « culte de soi californien », caractérisé par un « soi véritable » obscurci par de multiples influences et qu'il faudrait découvrir par le recours à la science, à un souci de soi antique qui repose plutôt sur un soi à construire (Foucault, 2015, p. 517). Parce que ce virage d'un soi introspectif vers un soi créatif nous semble pertinent dans notre démarche, nous proposons de nous arrêter quelques instants sur la conception foucauldienne du rapport à soi.

Pouvoir, domination et liberté

La critique sépare habituellement l'œuvre de Foucault en trois domaines d'étude qui se sont succédé au fil de la vie du philosophe : théorie des discours, analyse du pouvoir, étude du sujet. Jusqu'à la publication de ces derniers travaux au début des années 80, les recherches de Foucault furent interprétées comme une généalogie des systèmes de pouvoir occidentaux à travers l'étude de la discipline, de la norme, de la prison, du contrôle, du gouvernement, du libéralisme. Nous avons déjà introduit la notion de pouvoir et de gouvernementalité chez Foucault dans le premier chapitre. Pour pousser plus en avant leur exploitation dans le cadre ce chapitre, il nous semble intéressant de compléter ce premier compte-rendu par la distinction que Foucault opère entre relation de pouvoir et de domination.

Sur la question de la libération, Foucault critique la perspective selon laquelle il existerait une nature humaine libre qui aurait été mise sous verrou à travers une série de processus historiques et sociaux et qu'il faudrait retrouver. Interrogé sur l'existence d'un sujet souverain et libéré, Foucault donnera une réponse qu'il nous semble intéressant de retranscrire dans son intégralité :

J'ai toujours été un peu méfiant à l'égard du thème général de la libération dans la mesure où, si l'on ne le traite pas avec un certain nombre de précautions et à l'intérieur de certaines limites, il risque de renvoyer à l'idée qu'il existe une nature ou un fond humain

qui s'est trouvé, à la suite d'un certain nombre de processus historiques, économiques et sociaux, masqué, aliéné ou emprisonné dans des mécanismes, et par des mécanismes de répression. Dans cette hypothèse, il suffirait de faire sauter ces verrous répressifs pour que l'homme se réconcilie avec lui-même, retrouve sa nature ou reprenne contact avec son origine et restaure un rapport plein et positif à lui-même. Je crois que c'est là un thème qui ne peut pas être admis comme cela, sans examen²⁰. (Foucault, 1999b, p. 710)

Pourtant, il se défend tout autant de fournir une lecture fataliste de notre rapport au pouvoir selon laquelle nous serions emprisonnés et opprimés dans le système qui est le nôtre. Pour reprendre ses mots, une société « disciplinaire » n'est pas pour autant « disciplinée » (Foucault, 1999d, p. 15). Ces relations de pouvoir qui « aménagent la probabilité » (Foucault, 1999f, p. 237) sont fluides et mobiles. La liberté en est un élément fondamental. Pour que le pouvoir puisse s'exercer, il faut que chaque participant dispose d'un degré minimum de liberté lui permettant de déployer l'une ou l'autre forme de résistance. Le pouvoir n'est alors pas seulement répressif mais également productif. Toutefois, l'exercice de la violence nous fait sortir du domaine du pouvoir :

Une relation de violence agit sur un corps, sur des choses : elle force, elle plie, elle brise, elle détruit : elle referme toutes les possibilités ; elle n'a donc auprès d'elle d'autre pôle que celui de la passivité ; et si elle rencontre une résistance, elle n'a d'autre choix que d'entreprendre de la réduire. (Foucault, 1999f, p. 236)

Lorsque le champ de possibilité se réduit et que la détermination devient totale, la relation de pouvoir disparaît pour laisser place à un état de domination où il n'y a plus de fluidité ni de mouvement possible. La libération d'un état de domination est la précondition à l'exercice de pratiques de liberté. Pour mieux comprendre la distinction entre libération et pratique de liberté, Foucault donne l'exemple de la colonisation :

Je ne veux pas dire que la libération ou telle ou telle forme de libération n'existent pas : quand un peuple colonisé cherche à se libérer de son colonisateur, c'est bien une pratique

²⁰ Dans une démarche similaire, il déclarera également : « Je pense effectivement qu'il n'y pas de sujet souverain, fondateur d'une forme universelle de sujet qu'on pourrait retrouver partout. Je suis très sceptique et très hostile à cette conception du sujet. Je pense au contraire que le sujet se constitue à travers des pratiques d'assujettissement, ou, d'une façon plus autonome, à travers des pratiques de libération, de liberté, comme, dans l'Antiquité, à partir bien entendu, d'un certain nombre de règles, styles, conventions, qu'on retrouve dans le milieu culturel » (Foucault, 1999i, p. 733).

de libération, au sens strict. Mais on sait bien, dans ce cas d'ailleurs précis, que cette pratique de libération ne suffit pas à définir les pratiques de liberté qui seront ensuite nécessaires pour que ce peuple, cette société et ces individus puissent se définir des formes recevables et acceptables de leur existence ou de la société politique. C'est pourquoi j'insiste plutôt sur les pratiques de liberté que sur les processus de libération, qui, encore une fois, ont leur place, mais ne me paraissent pas pouvoir, à eux seuls, définir toutes les formes pratiques de liberté. (Foucault, 1999b, p. 710)

Foucault donne aussi l'exemple de la sexualité. S'il a fallu une étape première de libération pour se séparer d'une certaine domination d'une morale répressive, ce n'est pas pour autant qu'apparaît un sujet libre et épanoui. De nouveaux rapports de pouvoir viennent s'instaurer et nécessitent de nouvelles pratiques de liberté. Exercer sa liberté ne consiste pas à s'extraire des pressions externes, mais à les mettre en forme de façon consciente. A partir du début des années 80, le programme de recherche de Foucault dévie progressivement d'une analyse des dynamiques de pouvoir et de leurs influences sur les comportements à une analyse de ces pratiques de liberté. C'est dans ce cadre qu'il dirige son attention vers l'antiquité gréco-romaine et ses rapports à la sexualité.

Culture de soi

Pour comprendre comment Foucault en vient à proposer une éthique comme pratique de la liberté mais aussi comme démarche esthétique, il est nécessaire de s'intéresser à ses derniers travaux. Alors que dans la majorité de son œuvre, Foucault s'est intéressé à l'occident de la Renaissance et du 19^{ème} siècle, il dirige progressivement son attention à partir des années 70 vers l'antiquité gréco-romaine. Après la folie et l'emprisonnement, il se tourne cette fois vers la sexualité. Il ne réalise pas pour autant une histoire des pratiques sexuelles mais plutôt une histoire de la problématisation de la sexualité. Comment est-ce qu'à l'époque antique, un individu se définit comme le sujet moral d'une sexualité ? Comme le souligne Gros, « ce qui se trouve pris comme objet de recherche, ce sont donc des structures historiques de réflexivité, des manières historiquement déterminées et repérables de se rapporter à soi » (Gros, 2002, p. 232). Dans le prolongement de son analyse du gouvernement des autres, Foucault propose donc une analyse du gouvernement de soi, sur la façon dont les sujets se rapportent à eux-mêmes.

Foucault identifie dans la culture grecque du 4^{ème} siècle avant J.C. mais également dans la période impériale romaine des deux premiers siècles de notre ère, une phase déterminante dans notre

rapport éthique à nous-même. C'est à cette époque qu'il repère ce que les Grecs appellent *epimeleia heautou* et les Latins *cura sui*, ce qu'il traduit par le souci de soi (Foucault, 2015). Ce précepte est un des principes majeurs d'un art de l'existence qui s'étend sur plusieurs siècles et qui se retrouve dans la plupart des écoles de pensées majeures de l'époque :

« On sait bien que l'objectif principal des écoles philosophiques grecques ne consistait pas dans l'élaboration, l'enseignement d'une théorie. Le but des écoles philosophiques grecques était la transformation de l'individu. Le but de la philosophie grecque était de donner à l'individu la capacité qui lui permît de vivre différemment, mieux, de façon plus heureuse que les autres » (Foucault, 2013, p. 41).

Il repère la première élaboration de ce principe dans l'*Alcibiade* de Platon. A ce moment-là, le *souci de soi* traduit surtout un travail pédagogique sur soi de sorte à pouvoir exercer des fonctions de la cité et gouverner les autres. Foucault situe l'âge d'or de ce souci de soi dans la culture gréco-romaine des deux premiers siècles (Foucault, 2015). Il se transforme alors progressivement pour devenir un rapport durable à soi ainsi qu'une exigence qui s'applique à tous. Il ne s'agit plus de pallier à une éducation déficiente pour devenir un bon gouvernant mais de donner forme à son existence. Le rapport de soi à soi prôné par Sénèque, Plutarque ou Epictète vise à développer une maîtrise de soi-même. La connaissance de soi occupe une place primordiale dans cette relation critique à soi-même mais n'est pas la finalité. C'est dans le cadre du souci de soi que l'individu doit exercer une certaine forme de vigilance qui vise à vérifier que sa conduite générale soit en concordance avec ses pensées et principes. Enfin, la proximité lexicale de ce souci de soi avec le domaine de la médecine indique également une attention aux malaises qui circulent entre le corps et l'âme. De la même façon qu'il faut veiller à son corps, il faut également veiller à son âme. Cette pratique ne repose donc pas uniquement sur la reconnaissance d'un besoin d'instruction mais également la reconnaissance d'une souffrance qui doit être soignée par soi-même ou par quelqu'un ayant la compétence pour ce faire (Foucault, 1997).

L'ensemble de ces thématiques s'intègre dans une pratique volontaire dont le motif principal concerne l'élaboration de sa vie comme d'une œuvre d'art plutôt que d'un objet de connaissance (Foucault, 2015). Le soi n'est pas considéré comme quelque chose qui doit être découvert mais plutôt comme un objet en construction qui ne sera finalisé qu'à la fin de sa propre vie. L'objectif est l'élaboration d'une vie qui puisse servir d'exemple et dans laquelle on puisse se reconnaître. Ce

souci de soi vise également à l'élaboration d'une conduite appropriée qui permette de définir sa manière de se relationner aux autres ainsi que la portée de ses devoirs et obligations. C'est parce que ce souci de soi regroupe à la fois des pratiques sociales, des institutions, des techniques, des doctrines, des attitudes qui viennent s'articuler au sein d'une expérience à la fois individuelle et collective que Foucault parle d'une véritable « culture » de soi (Foucault, 2015).

Foucault constate ainsi que la différence majeure entre la morale gréco-romaine et le christianisme ne concerne pas la nature de leurs obligations et restrictions. Les deux courants de pensée reposent sur des thématiques ascétiques et la nature des interdits relatifs à la sexualité ne diffère pas énormément d'une période à une autre. Par contre, il constate que la façon dont l'individu compose sa subjectivité morale est complètement différente. C'est pour cette raison que Foucault propose de faire la distinction entre le code moral d'une part et l'éthique comme le processus à travers lequel on se constitue comme sujet moral d'autre part :

En somme, une action pour être dite « morale » ne doit pas se réduire à un acte ou à une série d'actes conformes à une règle, une loi ou une valeur. Toute action morale, c'est vrai, comporte un rapport au réel où elle s'effectue et un rapport au code auquel elle se réfère ; mais elle implique aussi un certain rapport à soi ; celui-ci n'est pas simplement « conscience de soi », mais constitution de soi comme un « sujet moral », dans laquelle l'individu circonscrit la part de lui-même qui constitue l'objet de cette pratique morale, définit sa position par rapport au précepte qu'il suit, se fixe un certain mode d'être qui vaudra comme accomplissement moral de lui-même ; et, pour ce faire, il agit sur lui-même, entreprend de se connaître, se contrôle, s'éprouve, se perfectionne, se transforme. (Foucault, 1999j, p. 558)

En fonction de l'école de pensée, morale et éthique évoluent avec plus ou moins d'autonomie. Les pratiques gréco-romaines se rapprochent plus d'une « morale orientée vers l'éthique » liées à des choix d'existence que les individus s'imposent eux-mêmes que d'une « morale orientée vers le code » (Foucault, 1999c). A l'inverse, les morales ultérieures seront plutôt tournées vers l'obéissance à une loi et un accomplissement éthique basé sur le renoncement de soi. C'est de cette façon que Foucault en vient à nouer liberté et éthique et à considérer « l'éthique du souci de soi comme pratique de la liberté » ou encore comme « la forme réfléchie que prend la liberté » (Foucault, 1999b, p. 712).

Techniques de soi

Foucault distingue plusieurs « formes » de techniques : les techniques de production, les techniques de signification, les techniques de domination et les techniques de soi. Les techniques de domination réfèrent aux enjeux de gouvernementalité identifiés précédemment. Le glissement opéré par Foucault vers le sujet se traduit par un intérêt croissant pour ces techniques de soi (Foucault, 1999g).

Le souci de soi antique ne se limite pas à une simple attitude générale à l'égard de soi. Comme le fait remarquer Foucault, le terme d'*epimeleia* implique un labeur qui s'inscrit dans le long terme et qui mobilise un large répertoire de techniques de soi. Foucault entend par là des « techniques qui permettent aux individus d'effectuer, seuls ou avec l'aide d'autres, un certain nombre d'opérations sur leur corps et leur âme, leurs pensées, leurs conduites, leur mode d'être ; de se transformer afin d'atteindre un certain état de bonheur, de pureté, de sagesse, de perfection ou d'immortalité » (Foucault, 1999g, p. 785).

Un art de la connaissance de soi se développe à travers des mises à l'épreuve (comme des exercices d'abstinence chez les épicuriens et les stoïciens), des examens de conscience et enfin un travail de la pensée sur elle-même :

« Veiller en permanence sur ses représentations, ou vérifier les marques comme on authentifie une monnaie, ce n'est pas s'interroger (comme on le fera plus tard dans la spiritualité chrétienne) sur l'origine profonde de l'idée qui vient; ce n'est pas essayer de déchiffrer un sens caché sous la représentation apparente ; c'est jauger la relation entre soi-même et ce qui est représenté, pour n'accepter dans le rapport à soi que ce qui peut dépendre du choix libre et raisonnable du sujet. » (Foucault, 1997, p. 68).

Cet examen de conscience vise à mettre à plat ce que l'individu a réalisé de sorte à pouvoir le confronter à un idéal de comportement qu'il se fixe. Foucault souligne que cette pratique se rapproche plus d'une logique de l'administrateur que du juge. Il s'agit de dresser un inventaire de ses succès et non de ses fautes morales. En fonction du constat de cet examen de conscience et de l'écart mesuré entre ce qui est fait et ce qui était désiré, l'individu révisé les lois qui formalisent sa conduite et décide s'il doit en réactiver certaines ou non. Concrètement, ces techniques de soi regroupent à la fois des exercices physiques, des soins du corps, des méditations.

Ces pratiques ne sont pas qu'individuelles mais s'inscrivent dans des structures institutionnalisées sous la forme d'entretiens avec un guide ou directeur. Travailler sur soi demande l'intervention d'un autre qui nous donne accès à des dimensions du soi qui ne seraient pas accessibles autrement. Cette pratique s'inscrit dès lors dans un jeu de devoirs et d'obligations réciproques qui contribuent à l'intensification des relations sociales. Se soucier de soi n'implique pas un désengagement mais plutôt l'élaboration d'un positionnement adéquat vis à vis des autres. C'est se demander quels sont ses devoirs en tant que père de famille, en tant que conseiller, en tant que dirigeant (Gros, 2002).

Si le Phèdre de Platon est connu pour sa critique de l'écriture à l'égard de son effet délétère sur la mémoire, Foucault constate toutefois que ce n'est qu'une considération particulièrement mineure en regard de l'ensemble du texte (Foucault, 2015). A l'inverse, l'écriture occupe une place toute particulière parmi les techniques de soi pour le travail sur la pensée qu'elle permet. Foucault (1999) distingue deux techniques particulières : l'écriture d'un *hypomnêmata* et la correspondance avec un ami intime.

L'hypomnêmata est un guide de conduite ou un livre de vie qui compose une mémoire de ce qui a été lu, entendu ou pensé. Son auteur y consigne des citations, des extraits de textes, des réflexions ou des raisonnements qui lui sont propres ou qu'il a entendus. Mais au-delà de son usage mémoriel, l'hypomnêmata sert à subjectiver les textes consignés. Foucault (1999) repère 3 caractéristiques qui permettent cette formation de soi : la stabilisation de l'écriture, une pratique du disparate et l'unification par l'appropriation.

Un élément récurrent dans la nature des techniques de soi est le recours aux autres, considérés comme des ressources essentielles. Toutefois, l'écriture est nécessaire pour stabiliser les matériaux récoltés et limiter la dispersion de sorte à constituer un capital sur lequel il est possible de revenir et de méditer. L'écriture permet de fixer l'attention qui risque sinon de se dissoudre en étant sans cesse tournée vers l'avenir. La fixation de ces matériaux ne doit pas pour autant conduire à la composition d'un corpus homogène qui délimiterait une école de pensée. Cela doit rester un ensemble hétérogène que l'individu peut réactiver selon les circonstances. Enfin, ce n'est pas pour autant qu'une forme d'unification n'est pas envisageable. À l'image d'un chœur qui s'impose à nous, mais dont ne peut pas dissocier chaque timbre, celle-ci prend la forme d'une appropriation de l'hétérogénéité qui doit pouvoir être reconnue à travers l'individu sans pour autant l'identifier.

La correspondance est une autre pratique de l'écriture de soi identifiée par Foucault. Cette lettre que l'on envoie à un ami proche peut viser à le conseiller, le rassurer ou le sermonner. Mais quel que soit l'objectif, il s'agit toujours d'un entraînement pour l'auteur qui, à travers les conseils qu'il produit, se prépare lui-même à devoir exercer les mêmes principes. Loin d'être unilatéral, cet échange évolue très vite vers une relation égalitaire où autant le maître que l'élève sont susceptibles de recevoir des conseils. Malgré les similarités entre la missive et l'hypomnêmata, la lettre est aussi une façon de se manifester à soi-même, une pratique de l'introspection à travers l'ouverture que l'on donne à l'autre sur soi. Cela peut passer dans un premier temps par un énoncé des interactions entre perceptions corporelles et pensées personnelles. On peut ainsi raconter comment une impression subjective, un malaise ou un problème de santé conduit à une méditation ou à un exercice de pensée. En se penchant sur les textes de Sénèque ou de Marc-Aurèle notamment, Foucault (1999b) constate que la lettre est également la version écrite de l'examen de conscience.

Souci de soi moderne

Foucault situe, plus ou moins à partir de Kant et l'époque de l'*Aufklärung*, que la philosophie s'éloigne d'une ontologie formelle de la connaissance pour se doter d'une nouvelle préoccupation qui est celle de son actualité et de la place de l'activité philosophique par rapport à cette actualité. Lorsqu'il réinterprète son parcours de recherche au début des années 80, Foucault déclare s'insérer dans cette préoccupation pour l'actualité à travers le projet général de ce qu'il a appelé une « ontologie historique de nous-mêmes » ou une « histoire critique de la pensée » (Foucault, 1999h). Lorsqu'il interroge la folie, la pratique de l'emprisonnement ou encore la sexualité, il s'intéresse à la façon dont ces objets se sont constitués comme des objets de connaissance à des moments précis au sein de communautés sociales particulières. De la sorte, il souhaite montrer en quoi certains objets qui se sont cristallisés comme des évidences sont en réalité des constructions socialement et culturellement déterminées pour, à terme, leur rendre leur mobilité conceptuelle. Lorsqu'il s'intéresse à la sexualité ou à la pratique de l'emprisonnement, ce n'est pas dans le cadre d'une histoire des comportements pour comprendre comment ils ont fait l'objet d'une problématisation, c'est-à-dire comme un « ensemble des pratiques discursives ou non discursives qui fait entrer quelque chose dans le jeu du vrai et du faux et le constitue comme objet pour la pensée (que ce soit sous la forme de la réflexion morale, de la connaissance scientifique, de l'analyse politique, etc.) » (Foucault, 1999e, p. 670). Comme le souligne Revel (2009), « l'histoire de la pensée problématise des objets, des modes de rapport à soi : elle s'interroge sur leur forme historiquement singulière et

sur la manière dont ils ont représenté à une époque donnée un certain type de réponse à un certain type de problème » (p.81). Par exemple, comment est-ce que la privation de liberté est devenue une manière raisonnable de punir ? Pour résumer avec les mots de Foucault :

Je cherche en effet à me situer à l'extérieur de la culture à laquelle nous appartenons, à en analyser les conditions formelles pour en faire la critique, non pas au sens où il s'agirait d'en réduire les valeurs, mais pour voir comment elle a pu effectivement se constituer. (Foucault, 1994)

Dès lors, son retour à l'Antiquité ne vise pas à faire l'apologie d'une sorte de « sagesse grecque » dans laquelle il trouverait une alternative viable aux modes de pensée modernes. Il cherche plutôt à provoquer un effet de décalage et à questionner nos évidences. Il insiste d'ailleurs bien sur le dégoût qu'il éprouve pour la morale grecque qui est avant tout une morale virile qui repose sur l'exclusion de l'autre et la dissymétrie (Foucault, 1999a).

Comme nous l'avons déjà souligné plus haut, il constate que ce souci de soi en tant que pratique éthique semble avoir disparu à l'heure actuelle. A sa place, on retrouve surtout la maxime socratique « connais-toi toi-même » ou encore les principes chrétiens d'ascétisme qui reposent sur un renoncement à soi. Cette survalorisation de la connaissance de soi est d'autant plus frappante que ce principe fût, pour une bonne partie de la philosophie antique, dans une relation de sujétion au souci de soi (Foucault, 1997). Se connaître est alors la première étape nécessaire au souci de soi. Foucault énonce plusieurs explications quant à ce renversement :

La première est le paradoxe éthique de l'ascétisme chrétien. Dans ce type d'ascétisme, le souci de soi doit prendre la forme d'un sacrifice : le renoncement à soi est l'objectif majeur du travail que nous devons réaliser sur nous-mêmes. La seconde raison est que la plupart de ces techniques de soi ont été intégrées, dans notre monde, dans des techniques éducationnelles et pédagogiques, médicales et psychologiques. Les techniques de soi ont été intégrées dans des structures d'autorité et de discipline, ou remplacées et transformées par l'opinion publique, les médias de masse, les techniques de sondage qui jouent un rôle formateur pour notre attitude à l'égard des autres et à l'égard de nous-mêmes, de sorte que la culture de soi est maintenant imposée aux gens par les autres et qu'elle a perdu son indépendance. La troisième raison est, je crois, que les sciences humaines supposent que la relation principale, majeure, à soi est, et doit être, essentiellement une relation de

connaissance. Et la quatrième et dernière raison est que, la plupart du temps, les gens croient que ce qu'on doit faire, c'est dévoiler, libérer, déterrer la réalité cachée du soi. (Foucault, 2015, p. 98)

Ce souci de soi que Foucault met en évidence ici n'est pas pour autant une « exception grecque ». La présence d'une culture de soi est, selon lui, un invariant culturel qui se retrouve à différentes époques. Toutefois, la visibilité et l'autonomie de cette culture de soi par rapport à un code moral est variable. Avec le christianisme, ce souci de soi se retrouve progressivement remplacé par un principe de connaissance de soi qui est mis au service d'un code moral. Se connaître n'est plus une étape d'un projet de stylisation de soi volontaire mais un processus d'introspection par lequel on vérifie que l'on respecte un code imposé. A partir de là, notre expérience de la morale se lie de plus en plus à différentes institutions religieuses, judiciaires ou scientifiques pour prendre la forme juridique d'un code à suivre (Foucault, 2015). Ce souci de soi n'a donc pas complètement disparu mais a perdu de son autonomie par rapport au code moral. Foucault ne repère que quelques résurgences occasionnelles de ce souci de soi à différents moments de l'histoire. Dans ces cas-là, cette préoccupation pour le soi prend la forme principale d'une libération à l'égard d'une moralité oppressive. Il remet toutefois en question la durabilité de ce rapport codifié à la morale :

Je crois que ces trois grandes références de notre éthique à la religion, à la loi et à la science sont maintenant, si je puis dire, usées. Et nous savons très bien que nous avons besoin d'une éthique, et que nous ne pouvons demander ni à la religion, ni à la loi, ni à la science de nous apporter cette éthique. Nous avons l'exemple d'une société, la société gréco-romaine, où une éthique [...] a existé sans ces trois références [...] Et c'est pour cette raison que, je crois, cette analyse de l'éthique gréco-romaine en tant qu'esthétique de l'existence peut avoir de l'intérêt. (Foucault, 2015, p. 143) ²¹

²¹ Il dira également : « Et si je me suis intéressé à l'Antiquité, c'est que, pour toute une série de raisons, l'idée d'une morale comme obéissance à un code de règles est en train, maintenant, de disparaître, a déjà disparu. Et à cette absence de morale répond, doit répondre une recherche qui est celle d'une esthétique de l'existence. » (Foucault, 1999i, p. 732). Ailleurs, il énoncera également : « Les mouvements de libération récents souffrent de ne pas trouver de principe sur lequel fonder l'élaboration d'une nouvelle morale. Ils ont besoin d'une morale, mais ils n'arrivent pas à trouver d'autre morale que celle qui se fonde sur une prétendue connaissance scientifique de ce qu'est le moi, le désir, l'inconscient, etc. » (Foucault, 1999a, p. 661).

Face à cette absence de repères moraux et faisant le constat qu'il n'y a pas de soi intérieur à retrouver, il propose l'élaboration d'une nouvelle pratique éthique de relation à soi-même qui repose sur une dynamique de stylisation de son existence similaire à ce que Foucault identifie dans l'Antiquité. Plutôt qu'un soi qui devrait être libéré ou actualisé, le soi doit être réalisé à travers la construction de relations de soi à soi. Plutôt qu'une essence à libérer, le soi est le produit de techniques historiquement situées, autant des techniques de pouvoir que de relation à soi. Ce n'est dès lors pas tant un soi aliéné qu'il faudrait libérer mais ces techniques mêmes qui sont à la racine de nos relations à nous-mêmes (Foucault *et al.*, 2017). Il faudrait élaborer de nouvelles formes de relations à soi. Ce n'est pas pour autant que le travail d'introspection est impossible. Si sa portée est bien limitée, cela ne l'empêche pas de servir de support à une pratique de construction de soi qui devient le marqueur d'une authenticité. Ce thème de l'esthétique de l'existence n'implique par un retour à un sujet souverain, autonome (pour reprendre ici son acceptation classique) mais plutôt que « le sujet se constitue à travers des pratiques d'assujettissement, ou, d'une façon plus autonome, à travers des pratiques de libération » (Foucault, 1999i, p. 733).

Cet appel à une esthétique de l'existence s'explique aussi par la relative hostilité qu'il éprouve à l'égard de la figure du philosophe comme donneur de leçon ou comme maître à penser :

Le rôle d'un intellectuel n'est pas de dire aux autres ce qu'ils ont à faire. De quel droit le ferait-il ? Et souvenez-vous de toutes les prophéties, promesses, injonctions et programmes que les intellectuels ont pu formuler au cours des deux derniers siècles et dont on a vu maintenant les effets. Le travail d'un intellectuel n'est pas de modeler la volonté politique des autres ; il est, par les analyses qu'il fait dans les domaines qui sont les siens, de réinterroger les évidences et les postulats, de secouer les habitudes, les manières de faire et de penser, de dissiper les familiarités admises, de reprendre la mesure des règles et des institutions et, à partir de cette problématisation (où il joue son métier spécifique d'intellectuel) de participer à la formation d'une volonté politique (où il a son rôle de citoyen à jouer). (Foucault, 1999e, pp.676-677)

Foucault fut critiqué à plusieurs reprises pour sa relecture « baudelairienne » de l'Antiquité qui, inspectée de plus près, se rapproche plus du dandysme que du souci de soi grec (Varga, 2009). A plusieurs reprises, Foucault fait lui-même le parallèle avec l'homme moderne de Baudelaire : « L'homme moderne, pour Baudelaire, n'est pas celui qui part à la découverte de lui-même, de ses

secrets et de sa vérité cachée ; il est celui qui cherche à s’inventer lui-même. Cette modernité ne libère pas l’homme en son être propre ; elle l’astreint à la tâche de s’élaborer lui-même » (Foucault, 1999h, p. 571).

Notre objectif n’est pas de rentrer dans les détails de l’analyse historique proposée par Foucault mais de l’utiliser selon son objectif initial, comme outil de décentrement. L’analyse de Foucault nous semble ici pertinente pour le modèle de construction du sujet qu’il propose et qui s’éloigne à la fois du modèle de la psychologie humaniste du sujet qui s’actualise par introspection et du modèle, plus récent, du gestionnaire qui optimise son rapport à soi (Gros, 2005).

Tout comme Verbeek (2011), nous identifions dans cette perspective foucauldienne de l’éthique une solution pour réconcilier la perte de souveraineté du sujet impliquée par une ontologie relationnelle avec l’idée d’un individu libre. Cette liberté ne s’exprime pas par l’absence de médiations mais par leur façonnage conscient. C’est en incorporant consciemment ces formes de médiation qu’un individu peut agir sur la constitution de sa subjectivité. Dans notre cas, nous avons préféré conserver la notion d’autonomie tout en limitant la portée de sa signification tandis que Verbeek préfère utiliser la notion de méta-agentivité.

Si Verbeek opère le même parcours que nous, il l’articule toutefois dans un projet de design tout à fait différent du nôtre. Cette reconnaissance du pouvoir moral de la technique ainsi que d’une responsabilité de méta-agentivité de la part de l’utilisateur conduit Verbeek à formuler une responsabilité morale chez le concepteur. Si la configuration d’un dispositif construit un mode d’existence, il devient possible de consciemment suggérer certaines façons d’agir plutôt que d’autres. Nous ne sommes plus sur une éthique qui s’intéresse à définir des dispositifs moralement acceptables, mais sur une éthique qui se préoccupe de *la bonne vie* (Verbeek, 2012). L’enjeu n’est donc plus de tracer la ligne de démarcation entre les technologies adéquates, mais de donner forme à nos relations avec ces dispositifs de sorte à conduire la société vers un idéal socialement partagé. Cette approche partage de nombreuses similitudes avec la méthode du *Value Sensitive Design* (B. Friedman, 1996) sur laquelle nous reviendrons plus loin. Contrairement à ces approches, notre travail de conception est plutôt dirigé vers le développement, chez l’utilisateur, d’une sensibilité à cette médiation technique de son rapport au monde. Plutôt que d’encourager un comportement spécifique, nous souhaitons soutenir la capacité individuelle à identifier comment un dispositif

technique peut pousser à l'adoption de représentations, d'attitudes ou de comportements particuliers.

Si elle nous permet de tracer une trajectoire générale, cette conceptualisation de l'autonomie n'est pas encore assez précise pour nous permettre de l'opérationnaliser dans une pratique de design. Toutefois, envisager l'autonomie comme une pratique esthétique de production de soi, comme double dynamique d'assujettissement et de résistance, reposant à la fois sur des pratiques de lecture, d'écriture, de méditation ou encore d'échange avec un maître de pensée nous ouvre plusieurs portes intéressantes à explorer pour approfondir notre approche :

- L'autonomie comme un processus dynamique de création socialisé ;
- L'autonomie comme un processus distribué ;
- L'autonomie comme une attitude critique.

DIMENSIONS DE L'AUTONOMIE

Autonomie comme compétence créative

La notion d'autonomie relationnelle apparaît aux alentours des années 2000 à travers une série de travaux féministes en philosophie morale. Ce terme ne désigne pas une notion clairement identifiable mais plutôt un faisceau de critiques qui visent à réhabiliter le concept en lui reconnaissant une dimension sociale et relationnelle (Mackenzie et Stoljar, 2000). Plutôt que de reposer sur une valorisation de l'autosuffisance et de l'indépendance rationnelle, ces approches postulent le sujet comme fondamentalement vulnérable et dépendant. La seconde prémisse de cette conception est une ontologie sociale du sujet, c'est à dire un sujet constitué par ses pratiques sociales, ses groupes d'appartenance et une série de contingences historiques. Plusieurs courants théoriques alimentent ces critiques et il n'est pas surprenant d'y retrouver à plusieurs reprises une perspective foucaldienne (Mackenzie et Stoljar, 2000) mais aussi une mention générale aux travaux en sciences cognitives qui étendent la cognition au-delà du système nerveux (Christman, 2009). Les approches de Friedman et Christman, discutées plus tôt dans ce chapitre, sont généralement associées à ce courant de recherche. Cette perspective entend appréhender l'autonomie dans toute sa complexité et son imperfection :

It [relational autonomy] does not assume that persons are ideally, or even hypothetically, fully rational agents or that their motivational structures are transparent to themselves.

Nor does it take as its starting point a model of an ideally just society and assess the extent to which current social relations fall short of this model. Rather, its starting point is the individual as situated in, shaped, and constrained by her sociorelational context in all its complexity; that is, its starting point is non-ideal agents in a non-ideal world, characterized by social oppression, injustice, and inequality. Given this starting point, the aim of relational autonomy theory is to theorize the kind of autonomy that is possible for non-ideal human agents; to diagnose how social domination, oppression, stigmatization, and injustice can thwart individual autonomy; and to hypothesize possible solutions, in the form of proposing how specific social relations, practices, and institutions might be reformed in such a way as to protect and foster individuals' autonomy. (Mackenzie, 2014, p. 23)

Nous avons précédemment souligné le rapport étroit, identifié par Foucault, entre le christianisme et l'appel moderne à l'introspection. Il nous semble que les travaux de Meyers qui formalisent l'autonomie comme un ensemble de compétences peut nous permettre de donner un peu de consistance à cet appel à la stylisation de notre existence.

La plupart des perspectives procédurales énumérées jusqu'à maintenant présentent l'influence des relations sociales sous un angle majoritairement négatif. La recherche de l'autonomie est conceptualisée comme une entreprise identitaire qui se réalise en confrontation avec ces influences (Mackenzie et Stoljar, 2000). Comme le fait remarquer Meyers, d'une façon qui n'est pas sans rappeler le reproche de Foucault à l'égard de la culture de soi moderne :

Since these theories take the true self off the autonomous individual to be a socialization-transcending agent, the procedures they commend for achieving autonomy are designed to cleanse the individual of the stain of socialization. Either these procedures are schemes to trick socialization into exposing itself so that its captive can escape and perhaps get the upper hand, or they are techniques whereby the cunning prisoner of socialization can break its stranglehold through sheer disdain for its power. (Diana T. Meyers, 1989, p. 43)

Pour expliquer comment la socialisation d'un individu peut à la fois contribuer à son oppression comme au développement de son autonomie, Meyers développe une conception de l'autonomie comme un répertoire de compétences décomposées en 3 catégories principales : découverte de soi,

définition de soi et direction de soi (1989). Plutôt que de rechercher un soi interne, cette perspective conduit à l'élaboration d'une identité dynamique.

Selon Meyers, dans une situation idéale, le processus de prise de décision repose sur une procédure formelle centrée sur une analyse rationnelle des éléments pertinents à la situation considérée et attentive aux conséquences. Toutefois, dans la majorité des cas, l'exercice de cette « rationalité délibérative » est impossible. Dans ce genre de situations, un travail d'introspection est nécessaire pour sonder ses sentiments, ses émotions, ses valeurs. Cette pratique de lecture de soi implique l'exercice de plusieurs compétences qui gravitent autour d'un travail d'identification de ses réactions intimes. L'exercice de compétences mémorielles ainsi que de son imagination contribuent à l'identification et à l'interprétation de ce que Meyers appelle des réponses autoréférentielles ou signaux affectifs. La communication permet de recourir à l'aide d'autrui pour disposer d'interprétations multiples à opposer à sa propre lecture. L'imagination sert également, une fois le processus interprétatif réalisé, à se projeter dans des situations potentielles permettant d'éprouver les hypothèses explicatives développées. Une fois encore, ce processus peut autant être solitaire que collectif. A terme, cette phase d'introspection que Meyers qualifie de découverte de soi, doit permettre à l'individu de se composer un « portrait » personnel. La composition de ce portrait est le fondement d'une démarche de définition et de direction de soi. La définition d'un plan de vie, pour reprendre l'expression de Meyers, démarre de ce travail introspectif qui permet à l'individu de définir une série de valeurs ou d'engagements qu'il souhaite tenir. De la même façon que pour le principe de découverte de soi, l'imagination et la communication contribuent à la confection d'une série d'options qui seront évaluées en fonction de leur désirabilité, elle-même fondée sur la proximité entre ces options et le portrait de soi qu'il aura dressé lors de sa phase d'introspection. Meyers définit, par direction de soi, la capacité d'un individu à assumer la responsabilité des jugements issus des deux phases antérieures et à les mettre en œuvre en situation réelle suffisamment longtemps pour pouvoir éprouver leur validité.

Bien qu'elles soient présentées comme des dimensions distinctes, l'exploration, la définition et la direction de soi s'entremêlent dans un mouvement perpétuel d'essais-erreurs qui permettent de reconfigurer dynamiquement l'identité de l'individu en fonction de son évolution personnelle :

The difference between autonomous people and nonautonomous ones depends on the capabilities people have at their disposal and the way in which people go about

fashioning their lives. Autonomous people are not vouchsafed a glimpse of their inner selves that other people are denied. Rather, they possess and exercise skills that maintain a fluid interaction between their traits, their feelings, their beliefs, their values, their extended plans, their current possibilities for realizing these plans, and their conduct. (Diana T. Meyers, 1989, p. 55)

A cet égard, l'expression de soi reste une tâche approximative qui demande une redéfinition constante et donc une capacité d'autocorrection. Toutefois, de façon similaire à Christman (2009), Meyers souligne que ce travail d'introspection n'est pas une tâche continue mais repose plutôt sur une attitude de veille à l'égard des besoins d'investigation qui peuvent se faire ressentir. Honte, dégoût de soi, embarras sont autant de signaux qui indiquent un besoin d'introspection et de révision.

Pour résumer, le répertoire de compétences proposé par Meyers implique à la fois des compétences de raisonnement et de communication mais également des compétences émotionnelles (interpréter et réguler ses émotions) ainsi que des compétences imaginatives pour pouvoir entrevoir les conséquences de ses actes et planifier des lignes de conduites alternatives. Appréhendée comme un registre de compétences à maîtriser, l'autonomie s'exprime en degré plutôt que selon une catégorisation binaire. L'implication de compétences sociales et émotionnelles implique également que l'exercice de l'autonomie est impossible sans socialisation préalable de l'individu. Toutefois, cela signifie également que certains schémas culturels ou structures sociales peuvent autant contribuer au développement de certaines de ces compétences qu'empêcher leur développement. Entre autre, l'éducation est pointée par Meyers comme un élément central du développement de ces compétences, en bien comme en mal.

Autonomie comme compétence distribuée

Penser l'autonomie comme un processus créatif plutôt que comme un retour vers un soi authentique permet d'appréhender plus aisément la nature distribuée de cette compétence. L'approche par compétence a pour avantage majeur de souligner la nature acquise de ce comportement ainsi que l'importance de conditions structurelles au développement d'un comportement autonome. Toutefois, dans ces approches, le rapport qu'entretient l'individu avec les conditions structurelles de son autonomie relève généralement de la passivité. L'action est plutôt attendue par une autorité extérieure (état, institution, communauté) pour garantir le maintien de conditions propices à

l'autonomie de l'individu. Toutefois, être autonome implique également d'être capable de moduler, dans la mesure de ce qui est possible, la configuration de son environnement social et technique de sorte à soutenir l'une ou l'autre dimension de sa volonté. Explorons deux exemples issus de la littérature en philosophie morale.

Le premier exemple est une anecdote personnelle présentée par Diana Tietjens Meyers (2005) dans un article intitulé « Decentralizing Autonomy ». Souffrant d'une condition médicale l'empêchant de métaboliser certains aliments, elle doit adopter certaines restrictions alimentaires. Naturellement, elle partage la nouvelle avec plusieurs connaissances. Après un certain temps, elle constate qu'en présence de connaissances au fait de sa condition, elle s'empêche de céder à certaines tentations qui lui seraient nuisibles. Elle n'avait évidemment pas prévu cette situation et le partage de cette information n'était pas motivé par une telle stratégie. Toutefois, après avoir constaté les faits, elle peut mobiliser cette caractéristique de son réseau social pour renforcer sa résolution et faciliter l'endossement de ses impératifs diététiques.

Un deuxième exemple notable est celui des communautés de choix présenté par M. Friedman (1989). Si certaines communautés nous sont imposées telles que notre famille ou notre culture, nous sommes parfois amenés à rejeter certaines des valeurs et des engagements qui sont promus par ces collectivités. Composer de nouvelles relations ou s'associer à des communautés par choix personnel peut contribuer à soutenir l'exercice de son autonomie et renforcer son adhérence à des normes spécifiques. Pour reprendre les termes de M. Friedman (1989):

Perhaps it is more illuminating to say that communities of choice foster not so much the constitution of subjects but their reconstitution. [...] The modern self may seek new communities whose norms and relationships stimulate and develop her identity and self-understanding more adequately than her unchosen community of origin, her original community of place. In case it is chosen communities which help us to define ourselves, the project of self-definition would not be arising from communities in which we merely found or discovered our immersion. It is likely that chosen communities, lesbian communities, for example, attract us in the first place because they appeal to features of ourselves which, though perhaps merely found or discovered, were inadequately or ambivalently sustained by our unchosen families, neighborhoods, schools, or churches.

Thus, unchosen communities are sometimes communities which we can, and should, leave, searching elsewhere for the resources to help us discern who we really are. (p.289)

Ces deux exemples nous amènent à souligner que le processus cognitif de réflexion critique discuté dans les sections précédentes ne porte pas uniquement sur la définition de lignes directrices générales guidant sa conduite mais également sur la position à adopter au sein d'un nexus d'influences et de pressions externes plus ou moins déterminantes :

Though people cannot choose directly to change their constitutive characteristics, they can choose to place themselves in situations and to act in ways designed to bring about such changes. [...] While it would be a mistake to insist that people can control all of their qualities, it is clear that people can indirectly choose to transmute at least some of their constitutive qualities and that they can contrive to alter many others to some extent. (Diana T. Meyers, 1989, p. 95)

Il nous semble que ce que Foucault décrit comme une relation esthétique élaborée à l'aide de techniques de soi partage de nombreuses similitudes avec une conception de l'autonomie comme compétence distribuée. Nous faisons ici volontairement le parallèle avec la notion de cognition distribuée discutée dans le premier chapitre. Dès lors, si cette autonomie peut être socialement distribuée, rien ne nous semble interdire de postuler une autonomie matériellement distribuée. C'est en tout cas ce que semble indiquer les études actuelles portant sur les usages d'écoute musicale en ligne via des plateformes de streaming (Beuscart *et al.*, 2019 ; Karakayali *et al.*, 2018). Leur autonomie ne s'exprime pas à travers leur refus de recourir à une guidance dont ils ne peuvent pas maîtriser l'ensemble des tenants et aboutissants. Elle s'exprime plutôt par leur capacité à identifier cette guidance et à décider s'ils souhaitent l'endosser ou non. Pour poursuivre le parallèle avec la théorie de la cognition distribuée, l'exercice de son autonomie implique une capacité de méta-autonomie. Quelles sont les influences, parmi celles que je peux plus ou moins contrôler, au sein desquelles je veux m'inscrire parce qu'elles m'assurent, avec plus ou moins de certitude, d'exercer des actes concordants avec mes désirs, mes valeurs ou mes engagements ? Cette perspective se retrouve notamment dans la conception de l'autonomie de Floridi *et al.* (2018), entendue comme la capacité d'un individu à céder volontairement la prise de décision à un dispositif intelligent. Autrement dit par Latour : « La question ne se pose plus de savoir si l'on doit être libre ou attaché, mais si l'on est bien ou mal attaché. [...] S'il ne s'agit plus d'opposer attachement et détachement,

mais les bons et les mauvais attachements, il n'y a qu'un seul moyen pour décider de la qualité de ces liaisons ; s'enquérir de ce qu'ils sont, de ce qu'ils font, apprendre à être affecté par eux. » (Latour, 2000, pp.3-4)

Certains auteurs en philosophie morale vont néanmoins plus loin et identifient des conditions structurelles qui sont nécessaires à l'exercice de son autonomie. Oshana identifie, par exemple, deux conditions propres aux conditions structurelles d'exercice de son autonomie : le contrôle et l'accès à un répertoire d'options pertinentes (Oshana, 2016). Un individu doit pouvoir disposer d'une capacité d'action ou d'un pouvoir suffisant pour empêcher d'autres personnes ou institutions d'exercer un contrôle coercitif sur lui. De plus, selon Oshana, agir de façon autonome implique d'avoir eu l'opportunité de confronter ses choix à un répertoire d'options alternatives viables de sorte à pouvoir exercer ce contrôle. L'absence d'opportunités risque de conduire à un raisonnement biaisé qui tente de justifier, a posteriori, un choix dont le facteur déterminant fut simplement d'avoir eu le mérite d'exister. A l'inverse, disposer d'alternatives viables implique de devoir refuser certaines options et donc de préciser ses préférences et ses besoins (Oshana, 2016). A ces conditions structurelles, Oshana ajoute encore une série de conditions d'indépendances substantives (notamment : disposer d'un niveau de sécurité suffisant, profiter d'une indépendance financière, pouvoir poursuivre des objectifs et des valeurs qui divergent de ceux poursuivis par des individus en position d'autorité, ne pas être victime de désinformation concernant ses capacités d'action réelles) (Oshana, 2016). Mackenzie rapproche cette préoccupation pour l'accès à des opportunités pertinentes à une conception positive de la liberté que l'on peut notamment retrouver dans la conception razienne de l'autonomie (Raz, 1986). Cette préoccupation pour des conditions structurelles nécessaires nous renvoie également à la distinction que Foucault opère entre domination et relation de pouvoir. Pour que des pratiques de liberté puissent s'épanouir, il est nécessaire que la relation garde en fluidité et ne s'enfonce pas dans une inertie unilatérale. Comme le souligne Verbeek (2011):

When technological mediations do not leave any room for « relational freedom » for human beings to constitute their (moral) subjectivity, they oppress and limit human subjects and must not be permitted to function as a basis for generating forms of subjectivity. (p.111)

Pour en revenir progressivement à des problématiques de design, il nous semble pertinent de rapprocher cette perspective de la notion d'autonomie issue du *value sensitive design* dans le domaine de l'interaction homme-machine. B. Friedman (1996) y définit l'autonomie comme une capacité d'auto-détermination : « Individuals who are self-determining, who are able to decide, plan and act in ways that they believe will help them to achieve their goals and promote their values » (p.17) Appliquée à la conception d'un programme informatique, une perception de l'autonomie comme liberté négative impliquerait de garantir à l'utilisateur l'accès à l'ensemble des fonctionnalités d'un programme. Toutefois, agir de la sorte conduira, dans de nombreux cas, à désorienter l'utilisateur qui, au final, perdra en autonomie d'action. Respecter l'autonomie de l'utilisateur implique donc de lui fournir juste assez d'accès et de fonctionnalités pour ne pas le surcharger tout en lui permettant de mener à bien les tâches qu'il souhaite réaliser. En prenant l'exemple d'un éditeur de texte, Friedman suggère que l'autonomie de l'utilisateur n'est pas seulement proportionnelle à la liberté d'action qui lui est permise. Un éditeur de texte disposant de fonctionnalités avancées peut diminuer l'autonomie d'un utilisateur ne disposant pas des compétences techniques pour s'approprier certaines fonctionnalités trop complexes. A l'inverse, une interface plus simplifiée opérant à un niveau d'abstraction plus élevé, bien que limitant les possibilités d'action de l'utilisateur, peut lui permettre de réaliser plus de choses. Préserver l'autonomie de l'utilisateur est alors conceptualisé comme la mise à disposition d'un degré de contrôle adéquat au bon moment. B. Friedman (1996) identifie 4 points d'attention nécessaires à la préservation de l'autonomie de l'utilisateur :

- Aptitudes du système (*system capability*) : Mettre à disposition des fonctionnalités permettant d'agir sur le système ;
- Complexité du système (*system complexity*) : Adapter la charge cognitive demandée par le système en fonction du profil d'utilisateur et des tâches réalisées ;
- Mauvaise représentation du système (*misrepresentation of the system*) : Éviter les informations fausses portant sur le fonctionnement du système ;
- Fluidité du système (*system fluidity*) : permettre à l'utilisateur d'adapter le système en fonction de l'évolution de ses besoins.

Autonomie comme attitude critique

En parallèle des enjeux des gouvernances de soi identifiés jusqu'à maintenant, Mackenzie insiste sur l'attitude d'auto-habilitation qui consiste à s'octroyer l'autorité normative ainsi que la responsabilité de ses décisions, valeurs et de l'orientation de sa vie en général.

Self-authorization involves, first, regarding oneself as responsible to oneself and as answerable and accountable to others, for one's beliefs, values, commitments, and reasons. [...] To be self-authorizing, a person must not only be capable of understanding and responding to this social demand but must also regard herself as a valid source of self-authorizing claims. (Mackenzie, 2014, p. 36)

Cela signifie être préparé à fournir des explications quant à son comportement et être prêt à prendre la responsabilité de ses valeurs et engagements, être également prêt à les défendre et à les réviser.

Mackenzie (2014) distingue trois sous-dimensions à cette attitude générale d'auto-habilitation :

- Respect de soi (*self respect*) : se considérer comme l'égal moral des autres et avoir le droit de développer une perspective personnelle ;
- Confiance en soi (*self-trust*) : avoir confiance en ses engagements, ses convictions, son jugement ;
- Estime de soi (*self-esteem*) : attribuer de la valeur à ses engagements, ses attitudes, ses convictions.

Pour qu'elle puisse être la source d'une pratique d'autocréation, il nous semble que cette attitude d'auto-habilitation doit également impliquer une certaine attitude à l'égard du monde. Elle se doit d'être une attitude critique, telle que Foucault l'identifie dans l'interprétation kantienne de l'*Aufklärung* et dans le dandysme baudelairien, qui permet de s'appréhender dans son actualité et d'imaginer autrement le présent (Foucault, 1999h).

AUTONOMIE ET SERVICE PUBLIC

Designer pour l'autonomie

Le trajet conceptuel que nous avons effectué autour de la notion d'autonomie nous a amené à la conceptualiser comme :

- 1) Une attitude critique qui vise à s'auto-légitimer comme auteur et objet d'un jugement à la fois normatif et créatif ;
- 2) Une pratique distribuée de création de soi, fondée sur des pratiques d'assujettissements et de résistances dynamiques ;
- 3) Un ensemble de compétences rationnelles, créatives et sociales.

En partant de là, nous proposons de fonder une pratique de design qui repose sur trois impératifs principaux interconnectés :

- **Agentivité** : Garantir à l'utilisateur une agentivité minimale sur le système à travers le contrôle et le paramétrage de l'algorithme ;
- **Légitimité** : Renforcer son sentiment de légitimité à l'égard de son action sur le système et de l'appropriation de son rapport au média. ;
- **Intelligibilité** : Rendre intelligible les effets de la relation de médiation selon 3 modalités
 - *Rétrospective* : Faciliter l'exercice d'introspection pour permettre à l'utilisateur de recontextualiser l'évolution de ses intentions et pratiques au cours du temps ;
 - *Prospective* : Faciliter l'exercice de projection dans le futur et de réflexion sur les alternatives possibles ;
 - *Introspective* : Permettre un temps d'arrêt pour faire le point entre ses pratiques passées et ses intentions futures.

Plusieurs modalités d'interaction homme-machine peuvent être envisagées pour répondre de façon simultanée à ces impératifs :

- **Contrôle** : Contrôler l'algorithme intervient au niveau de l'agentivité mais également de l'intelligibilité en permettant à l'utilisateur de matérialiser plusieurs configurations possibles et de visualiser leurs conséquences.
- **Visualisations et explications** : L'intégration de visualisations de données, d'explications ou de données chiffrées contribue à l'intelligibilité du rapport au média à travers la réduction d'une asymétrie informationnelle entre l'utilisateur et le système. Ces informations peuvent être de nature diachronique (comparaison entre plusieurs utilisateurs) ou synchroniques (évolution temporelle du parcours d'un seul utilisateur).

- **Interruption** : Perturber la fluidité d'usage d'une fonctionnalité peut forcer l'utilisateur à marquer un temps d'arrêt dans ses usages. Cette interruption peut concerner soit la production de recommandation, soit l'interruption d'une action de l'utilisateur. Cette technique, encore expérimentale, fut notamment employée par Twitter pour limiter des partages de contenus « impulsifs » en demandant à l'utilisateur de valider son action une deuxième fois.

Pour mieux comprendre la spécificité de notre approche, il nous semble important d'identifier en quoi elle se distingue de l'approche dominante en interaction homme-machine qui repose sur un principe d'actualisation de soi. Là où un design conçu pour l'actualisation de soi va principalement reposer sur une configuration adéquate de l'outil lors des premières connexions de l'utilisateur de sorte qu'il lui « colle à la peau », de sorte qu'il épouse ses « potentialités », qu'il réfrène certaines pulsions nuisibles et enclenche une dynamique vertueuse. A l'inverse, l'usage autonome d'un algorithme de recommandation n'implique pas de connaître des domaines de sa personnalité qui seraient préexistants à l'usage. Cela implique plutôt d'être attentif à la co-construction de son identité (ses goûts, valeurs, engagements etc.) en situation d'usage médiatisé par ces technologies. Cette pratique de l'autonomie comme souci de soi implique donc de porter son attention à la fois sur l'activité soumise à la médiation par le système technique ainsi que sur le système technique lui-même et à la façon dont il contribue à la configuration de cette identité. Être autonome dans son écoute musicale sur Spotify ou dans sa consommation cinématographique sur Netflix ne demande pas uniquement de comprendre comment les outils proposés par ces plateformes peuvent renforcer ou stimuler des goûts et des envies que l'individu juge trop ou trop peu développés. Cela implique également de s'interroger sur la façon dont ces médias reconfigurent ce qu'est l'écoute musicale ou ce qu'est le cinéma et en quoi ces systèmes font partie de dispositifs sociotechniques plus larges dont on souhaite intégrer les pratiques et les principes. Dès lors, un design conçu pour stimuler l'autonomie de l'utilisateur doit s'inscrire dans une logique ponctuelle de dérangement, de décentrement ou de désaxement. Elle doit encourager l'utilisateur à questionner ses pratiques, ses valeurs et ses intentions futures au fur et à mesure qu'il s'engage dans une pratique de création personnelle.

Modèle général

Avant de conclure, nous tenions à réintroduire notre conception de l'autonomie dans la problématique générale de conception d'un algorithme de service public. Dans un premier temps, nous avons insisté sur la prédominance, au sein de cette littérature, d'une préoccupation pour l'algorithme et pour les valeurs qu'il matérialise. Nous avons alors fait le constat d'une approche du service public reposant principalement sur une vision paternaliste du service public et un modèle tripartite « équité/responsabilité/transparence ». Nous avons notamment cherché à approfondir les enjeux relatifs à la notion de transparence qui nous apparaissaient encore trop peu opérationnelle pour être utile en l'état. Pour remplir ce vide conceptuel, la majorité du chapitre fut consacrée à la formalisation de la notion d'autonomie. Nous proposons maintenant de boucler la boucle en revenant aux enjeux de service public en formulant un nouveau modèle qui nous semble plus apte à guider les réflexions sur la conception d'un algorithme de recommandation au service du public.

Nous proposons de repartir de la distinction que Foucault opère entre code moral d'une part et pratiques de liberté d'autre part. Par principe, toute forme d'éditorialisation algorithmique risque de favoriser l'une ou l'autre catégorie de contenus médiatiques. Dès lors, nous avons souligné l'impossibilité apparente d'envisager un système qui soit entièrement dépourvu de biais. Bien que nous n'irons pas jusqu'à dire que la recherche ne peut parvenir à isoler des principes moraux de services publics sur lesquels rythmer la recommandation algorithmique, il apparaît toutefois qu'aucune mesure ne parviendra à établir un consensus clair. La prise en compte du développement de l'autonomie de l'utilisateur qui se retrouve de façon diffuse dans la littérature scientifique semble donc être le corollaire de toute conception d'un système de service public.

En complément de ces deux pôles majeurs, nous proposons de reprendre la notion d'explicabilité, plutôt que de transparence, comme variable d'ajustement. Floridi *et al.* (2018) proposent cette notion, au sein de leur synthèse des principes éthiques relatifs à l'intelligence artificielle, pour son double sens d'intelligibilité et de responsabilité (*accountability*). Un système explicable doit à la fois garantir la compréhension de son fonctionnement mais aussi permettre de vérifier qu'il réalise bien les objectifs annoncés (qu'ils soient exprimés en termes de bienveillance ou de non-malfaisance). Tintarev et Masthoff (2007) privilégie la notion de scrutabilité pour exprimer un enjeu similaire.

Sur base de cette formalisation de la notion d'autonomie ainsi que sur la revue de littérature des travaux en design de systèmes de recommandation, nous proposons de formaliser la notion d'algorithme de service public selon un graphique radar à 3 pôles : Editorialisation / Agentivité / Explicabilité.

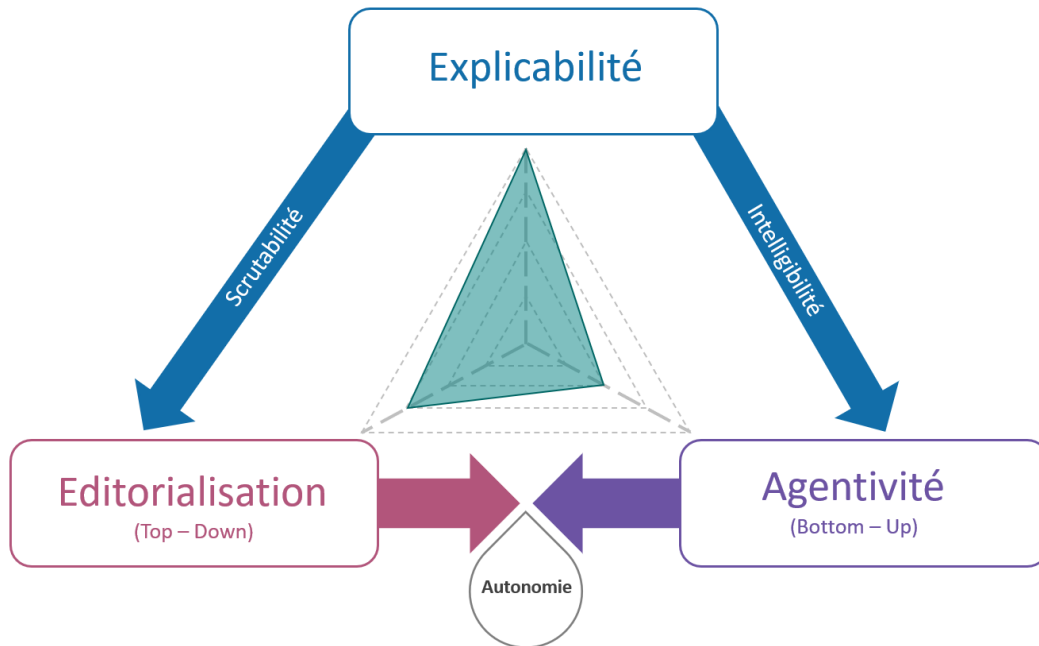


Figure 7 — Modélisation en 3 pôles d'un système de recommandation de service public

Le schéma en question n'est pas un graphique radar à proprement parler étant donné que les 3 variables sont plutôt qualitatives que quantitatives. Cette représentation graphique vise surtout à partager l'intuition qu'un système de recommandation pour un service public se développe dans l'interstice de ces 3 dimensions qui dessine un champ de possibilités. Par la notion d'éditorialisation, nous désignons la stratégie de publication et les valeurs qui sont intégrées dans le système algorithmique et qui se traduisent généralement par un paternalisme plus ou moins assumé. En fonction des missions de service public du média concerné, l'éditorialisation des contenus sera plus ou moins marquée par une certaine conception du service public. Cette conception est dépendante du contexte politique, économique et médiatique de la communauté

concernée et des préoccupations qui émergent de cet écosystème. Un hypothétique média de service public américain caractérisé par un écosystème médiatique et politique particulièrement polarisé n'aura pas les mêmes objectifs qu'un média de service public européen évoluant dans un paradigme très différent. La dimension de l'agentivité réfère aux opportunités d'actions sur le système qui sont mises à disposition de l'utilisateur.

L'explicabilité du système sert de variable d'ajustement entre la place laissée à l'éditorialisation et l'agentivité. Elle rentre en interaction avec ces deux pôles selon deux logiques différentes. L'explicabilité orientée vers l'activité d'éditorialisation concerne un devoir de scrutabilité (Tintarev et Masthoff, 2007) du média de service public. Cette scrutabilité désigne la présence d'outils permettant au média, à un organisme externe ou à ses utilisateurs d'évaluer dans quelle mesure les comportements de l'algorithme correspondent aux objectifs normatifs qui furent initialement fixés. La dynamique d'intelligibilité désigne plutôt les outils qui visent à permettre à l'utilisateur de comprendre l'effet de la médiation technique sur ses comportements. Ces deux facettes ne sont pas mutuellement exclusives et peuvent cohabiter au sein d'une même interface.

Là où les deux démarches diffèrent concerne la focale de l'explication. Dans le cas de la scrutabilité, le point d'attention est global (quel est l'effet de l'outil sur la communauté d'utilisateurs ?) tandis que la démarche d'intelligibilité démarre de l'individu (quelle est l'influence de cet outil sur ma pratique personnelle ?). Au-delà d'une intelligibilité du système technique, il s'agit de l'intelligibilité de la relation qui s'établit entre le dispositif et l'utilisateur.

La notion d'autonomie à laquelle nous faisons référence fut détaillée ci-dessus. Elle découle de la capacité de l'utilisateur à s'approprier, d'une part, l'éditorialisation imposée par l'algorithme et, d'autre part, les possibilités d'action sur le système. Elle est dès lors dépendante des dynamiques d'intelligibilité et de scrutabilité.

Enfin, avant de conclure, il nous semble important de souligner la normativité propre à cette notion d'autonomie. Nous avons motivé cette «quête» de l'autonomie par la préexistence de préoccupations nationales ou internationales à l'égard du développement d'un rapport critique à l'information. Malgré la dimension socio-relationnelle que nous avons adoptée, la notion d'autonomie reste marquée par un profond héritage occidental. Nous ne doutons pas que d'autres impératifs critiques soient possibles. Dès lors, ce modèle que nous proposons est probablement incomplet ou biaisé. Plutôt que la modélisation d'une réalité incontestable, il est avant tout un outil

de réflexion pour penser l'articulation entre les 3 dimensions proposées. A l'heure actuelle, ce modèle circule au sein de la RTBF comme outil de travail entre des employés de départements différents avec une fonction qui n'est pas sans rappeler celle d'*objet frontière* (*boundary object*). Pour reprendre la définition de Star et Griesemer (1989):

Boundary objects are objects which are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use. They may be abstract or concrete. They have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable, a means of translation. (p.393)

Ce n'est ici qu'une hypothèse et il nous faudra suivre l'évolution de ce modèle pour appréhender comment il cristallise, ou non, certains débats entre des profils techniques et administratifs au sein de la RTBF.

CONCLUSION INTERMÉDIAIRE

Le premier chapitre de cette thèse fut consacré à l'exploration des enjeux sociaux relatifs à l'exercice d'un pouvoir algorithmique. A travers la notion d'affordance, nous nous sommes attachés à décrire la forme que peut prendre une relation de pouvoir qui s'établit entre un dispositif algorithmique et ses usagers. Cette question du pouvoir algorithmique apparaît comme centrale au sein des débats qui interrogent la possibilité d'un algorithme de recommandation de service public. Toutefois, cette focale sur l'éditorialisation opérée par l'algorithme occulte tout un domaine de réflexion encore naissant portant sur l'influence de l'interface sur les usages de ces dispositifs.

La notion d'autonomie apparaît de façon récurrente dans la littérature scientifique comme contrepoids de l'algorithme. Toutefois, son usage multidisciplinaire se traduit par une multiplicité conceptuelle qui empêche la réalisation d'un travail fondamental sur le sujet. L'objectif premier du deuxième chapitre fut donc d'identifier ces différentes notions pour les resituer dans un paysage conceptuel plus large. Il est apparu que l'approche la plus développée à l'heure actuelle, mobilisée au sein du domaine des systèmes de recommandation, repose sur la notion d'actualisation de soi issue des théories psychologiques humanistes. Cette approche est sensible aux mêmes critiques que

son inspiration théorique, à savoir qu'elle n'est pas capable de décrire de façon adéquate la nature socialement constituée de notre identité ainsi que le caractère distribué de notre agentivité.

Dès lors, c'est en mobilisant les travaux de Foucault sur le sujet ainsi que les travaux plus récents de la philosophie morale, que nous avons tenté de conceptualiser l'autonomie de l'utilisateur tout en préservant ces fondamentaux. Nous en sommes arrivés à concevoir l'autonomie, non pas comme une pratique d'actualisation d'un soi authentique mais comme une pratique d'élaboration esthétique et critique. Cette conceptualisation distribuée implique d'associer une dimension rationnelle à des compétences d'imagination mais aussi de communication interpersonnelle.

De là, il nous a été possible de décrire brièvement quelques pistes de design qui pourraient conduire au développement de l'autonomie de l'utilisateur à travers le soutien de son agentivité, de son sentiment de légitimité et de l'intelligibilité de la relation au dispositif. Le prochain chapitre se veut dès lors méthodologique. Il sera dans un premier temps consacré à l'exploration de dispositifs expérimentaux mobilisant des modalités d'interaction homme-machine similaires à celles identifiées dans ce chapitre. Bien que les objectifs de ces travaux divergent souvent, ils nous permettront de constituer une première base technique sur laquelle construire un prototype expérimental. Dans un deuxième temps, nous présenterons en détail les fonctionnalités de ce prototype, qu'il s'agisse des fonctionnalités orientées utilisateur ou des fonctionnalités associées à la gestion de ce prototype. Enfin dans un troisième temps, nous présenterons le protocole méthodologique qui fut développé pour exploiter cet outil expérimental.

CHAPITRE 3 — MÉTHODOLOGIE

Analyser l’influence de configurations spécifiques d’un système de recommandation sur les usages de ses utilisateurs n’est pas chose aisée. Afin de situer notre approche méthodologique dans un paysage plus large, le premier objectif de ce chapitre consistera en une revue des méthodes classiques d’analyse de pratiques médiatiques en ligne issues des sciences sociales. Nous présenterons notamment les avantages et inconvénients de notre protocole qui repose sur le développement d’une plateforme expérimentale.

Une fois ce travail introductif réalisé, nous pourrons reprendre là où nous nous sommes arrêtés lors de notre chapitre précédent. A la suite des premières réflexions de design qui y furent rapidement abordées, nous allons maintenant tenter d’identifier de façon plus détaillée les modalités d’interaction homme-machine qui sont plus à même de soutenir les 3 dimensions de l’autonomie (agentivité, légitimité, intelligibilité) issues de notre modèle. Bien que la littérature relative à l’expérience utilisateur de systèmes de recommandation se concentre en grande partie sur la satisfaction immédiate de l’utilisateur ou sur des enjeux de transparence²², les travaux existants nous offrent un riche répertoire d’outils qu’il nous est possible d’articuler pour atteindre nos objectifs. En faisant l’état des lieux des quelques prototypes existants, nous concentrerons principalement notre analyse sur les fonctionnalités permettant de donner à l’utilisateur une forme de contrôle sur le système.

Une fois ce travail théorique terminé, nous présenterons *Alveho*, le prototype qui fut réalisé dans le cadre de nos recherches et qui s’inspire à plusieurs niveaux des études préexistantes dans le domaine. En plus de présenter les fonctionnalités de l’interface utilisateur, nous détaillerons également les fonctionnalités de gestion de la plateforme dont la genèse ainsi que le fonctionnement actuel sont intimement liés à la méthode de recherche. Enfin, nous conclurons ce chapitre en présentant le protocole méthodologique qui fut élaboré pour réaliser une première étude expérimentale sur base de cet outil.

²² Pour une critique de cette notion de transparence, nous renvoyons le lecteur au deuxième chapitre et plus précisément, à la section dédiée à la notion d’algorithme de service public.

MÉTHODOLOGIE GÉNÉRALE

Préambule méthodologique

Il est généralement convenu, selon différents manuels prescriptifs relatifs à la réalisation d'un projet de recherche, que la définition du protocole méthodologique ne se fasse qu'après un premier travail théorique de problématisation et de formalisation d'une question de recherche. En ce qui nous concerne, bien que ces éléments n'arrivent qu'au troisième chapitre de cette thèse, la réflexion quant au protocole méthodologique fut initiée avant tout exercice théorique. Il nous a d'abord semblé primordial, bien avant d'interroger notre objet de recherche que sont les systèmes de recommandation, de faire un état des lieux préalable de ce qu'il est possible de savoir sur ces dispositifs et des méthodes d'enquêtes existantes. En bref, qu'est-il possible d'élaborer comme connaissance scientifique sur un écosystème complexe qui comprend, entre autres, des plateformes de diffusion, des réseaux sociaux, des services publics mais également des communautés d'utilisateur, des logiques de marchés, des pratiques de production de soi, des dynamiques de surveillance, etc.

Étudier la dimension technique de ces dispositifs est particulièrement difficile en dehors de toute collaboration étroite avec un acteur du milieu. Si certains articles scientifiques ou brevets sont accessibles publiquement, il est dans l'ensemble difficile d'avoir une compréhension précise des mécanismes à l'œuvre. Ce constat fragilise dès le départ toute tentative scientifique qui souhaiterait rapprocher des comportements d'utilisateurs à des configurations techniques particulières. Pour compenser cela, il est possible, dans une certaine mesure, de pratiquer une rétro-ingénierie qui vise à déduire des logiques de fonctionnement du système sur base de ses réactions à des inputs contrôlés. C'est notamment ce que fait Roth *et al.* (2020) en explorant de façon automatisée le réseau de recommandation qui s'étend autour d'un échantillon de vidéos sur Youtube. Hargreaves *et al.* (2018) se sont, quant à eux, intéressés à l'influence de l'algorithme de Facebook sur la visibilité de contenus politique en observant les contenus recommandés à de faux comptes dont les abonnements respectifs traduisent une orientation politique différente. Ce genre de méthodologie a pour inconvénient de reposer sur la génération de comportements artificiels qui sont censés simuler, avec plus ou moins de fidélité, des comportements réels d'utilisateurs. Elles ont néanmoins le mérite de décrire le paysage médiatique qui se présente à l'utilisateur tout en permettant l'analyse de corpus de données importants. Une alternative consiste à observer le comportement

d'utilisateurs réels. Cela peut notamment impliquer la réalisation d'entretiens avec des utilisateurs d'un dispositif particulier et de leur demander de réaliser des tâches spécifiques tout en s'imaginant en situation d'usage quotidien (voir par exemple Wiard *et al.*, 2022). Ces analyses qualitatives permettent d'observer des comportements plus nuancés et complexes qui dépassent le simple suivi naïf des recommandations algorithmiques.

Un autre paradigme méthodologique dominant repose sur l'analyse des traces de navigation des utilisateurs. L'accès à ces données peut soit se faire via l'établissement d'un partenariat direct avec l'entreprise responsable du service étudié soit par l'utilisation d'une API publique. Le premier cas est assez rare à l'exception de plateformes de streaming musical qui multiplient les collaborations scientifiques (voir par exemple Anderson *et al.*, 2020 ; Beuscart *et al.*, 2019). Le second cas repose sur l'exploitation d'un accès direct aux serveurs de l'entreprise qui fixe alors les conditions d'accès et les types de données qu'il est possible de récolter. Bien souvent, c'est le système économique du service qui va déterminer l'ouverture plus ou moins importante de cette API. L'utilisation de ces APIs pour la recherche consiste bien souvent en un détournement de leur fonction initiale qui est plutôt pensée pour des annonceurs ou d'autres initiatives commerciales qui, à terme, bénéficient à la plateforme. Cela implique que les données fournies sont structurées selon les cadres conceptuels de l'entreprise à travers la forme qu'elle donne à son API. L'analyse de ces données ne peut donc pas se passer d'une déconstruction soigneuse de ces cadrages imposés. Cette déconstruction passe par la documentation de la complétude des données (ce qui est rendu accessible et ce qui ne l'est pas), le niveau de détails (qui peut varier d'un point de données à un autre) ainsi que les évolutions de la plateforme au cours de l'étude (Rieder *et al.*, 2015). Une alternative notable à ces APIs implique de recourir au « scraping ». Cette technique consiste à générer dans un navigateur une page du site qui nous intéresse et de récupérer dans la structure de la page les informations recherchées. Autant la génération de la page que l'analyse et la récupération de son contenu peuvent être réalisées de façon automatisée. Le recours à cette technique demande de rester attentif aux évolutions de l'interface ce qui implique une adaptation constante des outils de récolte. De plus, cette pratique est souvent proscrite par les conditions générales d'utilisation de ces plateformes.

Dans tous les cas, quelle que soit la méthode d'accès aux données, le recours à ces techniques demande une certaine persévérance et relève souvent du bricolage méthodologique (Le Béhec et

Alloing, 2018). De plus, même si le chercheur parvient à conserver une liberté suffisante dans la structuration des données, celles-ci ne donnent pas pour autant un accès « immédiat » au réel. Si elles permettent de définir des tendances générales au sein de communautés spécifiques, le recours à des méthodes qualitatives demeure primordial pour assurer une compréhension fine des usages qui sont à l'écoute du contexte dans lequel ils s'incarnent. A cet égard, nous privilégions une approche qui s'apparente à la notion de *thick data* proposée par Latzko-Toth *et al.* (2016). Même s'il est techniquement possible de récolter une quantité importante de données sur une grande diversité de dimensions, il est parfois plus bénéfique de limiter la variété et le volume des données pour permettre de les « épaissir » à l'aide de récoltes qualitatives plus à même de fournir les détails nécessaires à la compréhension d'usages médiatiques parfois complexes.

Quoi qu'il en soit, l'opacité des services commerciaux existants limite la profondeur de l'analyse qui doit souvent se limiter à des observations de surface qu'il n'est pas toujours possible de connecter à des caractéristiques techniques précises. Or, comme nous avons pu le mettre en évidence dans le premier chapitre, il existe un écart important entre d'une part les techniques existantes dans la littérature et ce qui est réellement implémenté. Enfin, limiter l'analyse à des dispositifs existants empêche toute recherche de nature prospective.

Étant donné ces contraintes d'accès, nous avons préféré nous distancier de ces plateformes pour privilégier une méthodologie expérimentale reposant sur le développement d'un prototype qui est ensuite manipulé par les participants de l'étude. En agissant de la sorte, nous pouvons conserver un plus grand contrôle sur notre cadre méthodologique tout en connaissant précisément la nature du système technique qui est exploité. Cette méthodologie, issue des sciences de l'interaction homme-machine et habituellement dédiée à l'évaluation de l'expérience utilisateur, a pour inconvénient majeur de placer l'utilisateur dans des conditions artificielles de navigation. La phase d'observation se réalise généralement en laboratoire alors que les usagers n'en sont qu'à leur première rencontre avec l'artefact. Cela implique que ce type de méthode est plus approprié pour évaluer l'utilisation qui est faite d'un système technique plutôt que pour observer des usages développés sur un temps plus long dans un contexte social particulier. La distinction utilisation, usage et pratique est sujette à débat et il n'est pas notre objectif de rentrer dans ces considérations terminologiques (à ce sujet, voir Tilleul, 2020). Dans le cadre de cette recherche, nous distinguerons l'utilisation qui réfère au processus de manipulation d'un objet technique de l'usage

qui englobe plus largement « le contexte social et culturel dans et par lequel les technologies sont mobilisées » (Denouël, 2012, p. 68).

En ce qui concerne notre recherche, l'étude de la capacité d'autonomie de nos participants ne s'observe pas à l'échelle micro de la manipulation technique mais plutôt à une échelle plus large qui doit pouvoir prendre en compte la façon dont l'utilisateur sort du cadre suggéré par le dispositif et l'intègre dans son écosystème médiatique, à la fois informationnel, social et technique. Le défi méthodologique qui nous attend implique donc d'adapter une méthode habituellement consacrée à l'évaluation de la manipulation formelle d'un outil de sorte à pouvoir étudier son appropriation par nos participants. Ces considérations seront abordées dans la troisième section de ce chapitre, une fois que nous aurons posé la problématique de recherche et présenté notre prototype.

Problématique de recherche

A travers les deux premiers chapitres de cette thèse, nous avons dégagé les grandes lignes directrices d'un projet de recherche général qui vise à mieux comprendre comment certaines modalités d'interaction homme-machine d'un dispositif de recommandation automatisé influencent les usages de ses usagers. L'objectif final de ce projet vise à émettre des recommandations de design permettant le développement d'un système de recommandation de service public. Étant donné la richesse de cette thématique, nous avons opéré un premier rétrécissement en nous focalisant sur la question de l'autonomie en affirmant la nature fondamentale de cette compétence pour le développement d'un usage critique. Pour pouvoir mener à bien cette recherche, nous avons décidé d'opérer un deuxième rétrécissement en limitant notre analyse aux outils permettant à l'utilisateur de contrôler l'algorithme. Ce choix s'explique par le contraste entre l'engouement scientifique pour cette fonctionnalité et l'absence de données empiriques fiables. En effet, la littérature scientifique tend à donner une priorité importance à ce genre de fonctionnalités dans le cadre du développement d'un algorithme plus éthique, quelle que soit la signification accordée à ce terme. Or, assez peu d'études s'attardent à mesurer la réelle efficacité de ces fonctionnalités ainsi qu'à évaluer les facteurs personnels ou environnementaux qui pourraient influencer sur leur utilisation. Dès lors, la problématique de recherche qui sera évaluée dans le cadre de l'étude présentée dans ce chapitre peut être formulée de la façon suivante : *Quels sont les effets de la contrôlabilité d'un système de recommandation sur le développement de l'autonomie de ses usagers ?*

Même si la contrôlabilité est la focale principale de l'analyse, nous avons également introduit d'autres fonctionnalités (explications, bulles informatives, visualisations de données, statistiques) de sorte à pouvoir évaluer comment ces outils entrent en interaction (ou non) avec les outils de contrôle. La présence de ces outils complémentaires permettra également de préparer le terrain pour des études futures qui seront orientées plus spécifiquement sur ces mécanismes.

Bien que nous sommes restés ouverts à des observations qui dépassent le cadre de nos attentes, plusieurs hypothèses sous-tendent notre méthodologie et le développement de la plateforme expérimentale détaillée dans la section suivante. Celles-ci découlent des 3 dimensions de l'autonomie identifiées dans le chapitre précédent (agentivité, légitimité, intelligibilité) :

- H1 : La présence de fonctionnalités de contrôle de l'algorithme encourage l'utilisateur à interagir avec le système.
- H2: Contrôler le fonctionnement de l'algorithme de recommandation facilite la compréhension du fonctionnement du système
- H3 : Comprendre le fonctionnement du système encourage l'utilisateur à développer un jugement critique sur les effets de la recommandation sur ses pratiques
- H4 : Contrôler le fonctionnement de l'algorithme de recommandation encourage l'exploration de recommandations alternatives
- H5 : L'exploration de recommandations alternatives encourage l'utilisateur à développer un jugement critique sur les effets de la recommandation sur ses pratiques.
- H6 : L'exploration de recommandations alternatives encourage l'utilisateur à développer une stratégie d'intégration de la recommandation automatisée dans ses pratiques

Bien que nous ne l'ayons pas précisé pour des raisons de lisibilité, chacune des hypothèses directement liées à la question de la contrôlabilité peut être décomposée en 3 sous-hypothèses en fonction du type de contrôle considéré (input, output, process).

Structure générale du protocole

L'inconvénient majeur derrière la conception d'une plateforme comme Alveho réside dans le temps de développement conséquent qu'un tel projet implique. Il s'agit donc d'un investissement à long terme qui vise à soutenir plusieurs cycles d'expérimentations. La structure interne de l'application a été pensée pour faciliter la réutilisation de l'outil. Dans le cadre de cette section, nous détaillerons

le protocole méthodologique que nous avons mis en place pour réaliser une première expérimentation avec Alveho. Outre les objectifs de recherche inhérents à cette étude, nous cherchons également à évaluer les opportunités et les limitations méthodologiques de ce genre d'outils. Nous évaluerons plus en détail l'efficacité de cette méthode dans le chapitre 4.

Le schéma ci-dessous décrit la structure générale de notre protocole de recherche, en partant de la formulation de la problématique de recherche, en passant par le design du prototype et en concluant sur la réalisation d'une expérimentation censée répondre à notre problématique. Cette structure connaît trois boucles rétroactives :

- 1) La première boucle concerne la révision de la problématique de recherche en fonction de l'état de l'art des recherches et des prototypes existants.
- 2) La deuxième boucle concerne la phase de prototypage. Lorsqu'une première version d'Alveho fut produite, elle fut testée auprès de connaissances afin d'évaluer le fonctionnement des outils implémentés et de corriger d'éventuelles défaillances techniques. A la fin de cette phase de test, une deuxième phase de développement fut entamée afin de finaliser le prototype. Une deuxième boucle concerne la phase d'expérimentation. Le développement du prototype alterne entre des phases de design ou de re-design de fonctionnalités avec des périodes effectives de développement informatique.
- 3) En fonction des résultats de l'étude expérimentale, une nouvelle phase de développement peut être initiée afin d'explorer de nouvelles fonctionnalités ou de corriger des problèmes rencontrés lors de l'étude. En ce qui nous concerne, nous n'avons pas eu l'opportunité d'initier une nouvelle phase de développement pouvant déboucher sur une deuxième expérimentation. Toutefois, plusieurs perspectives de recherche issues de nos résultats sont identifiées à la suite de la présentation de nos résultats dans le quatrième chapitre.
- 4) Les résultats de l'étude peuvent également conduire à remonter plus loin dans le processus de recherche et à réviser la thématique de recherche avant d'initier une nouvelle phase de développement.

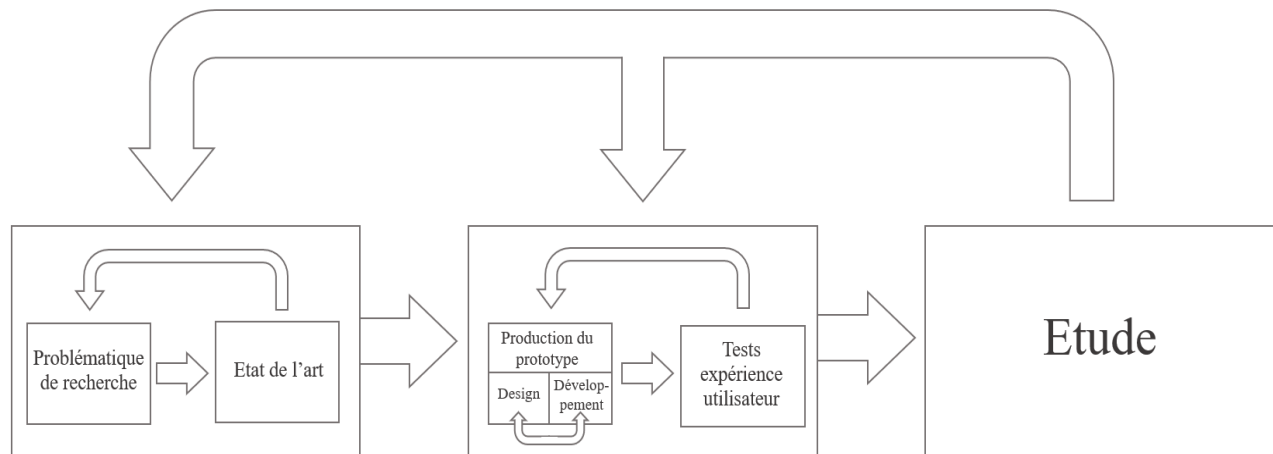


Figure 8 — Structure générale du protocole de recherche

PROTOCOLE DE CONCEPTION

Travaux préexistants sur la contrôlabilité

Si de nombreux travaux inspirés des recherches de Foucault se sont intéressés à l'influence de médias numériques sur le développement de capacités d'introspection, ces recherches se concentrent principalement sur des processus de découverte de soi à travers les autres via des outils de communication en ligne (blogs, réseaux sociaux etc.) (voir par exemple : Abbas et Dervin, 2009 ; Bakardjieva et Gaden, 2012 ; Haider, 2016). Nous n'avons pas connaissance de recherches qui se soient intéressées à cette thématique des technologies de soi dans le cadre d'un travail de design. Si Karakayali *et al.* (2018) reprennent bien cette thématique en s'intéressant aux pratiques de consommation musicale sur une plateforme de streaming, la seule dimension de contrôlabilité analysée concerne la possibilité offerte aux utilisateurs de sélectionner les musiques qu'ils souhaitent ajouter à leur profil. Dès lors, la majorité des travaux qui sont présentés ci-dessous découlent d'une démarche qui diffère sensiblement de la nôtre. Toutefois, même s'ils ne cherchent pas à mesurer le développement de compétences réflexives, ces travaux nous permettent d'identifier plusieurs mécanismes de contrôles d'un algorithme de recommandation. Il est ainsi possible de distinguer, de façon très grossière, un premier groupe d'initiatives issues des « sciences sociales » et un deuxième groupe issu du domaine de l'interaction homme-machine et des sciences informatiques.

Ce premier groupe d'initiatives se caractérise par des prototypes accessibles au public et par des préoccupations issues du domaine des sciences sociales. Ces initiatives, généralement bien médiatisées, ne débouchent pas nécessairement sur des publications scientifiques. Toutefois, les prototypes sont souvent accessibles en accès libre sur internet. *Gobo* est un projet du Center for Civic Media du MIT Media Lab, dirigé à l'époque par Ethan Zuckerman, qui vise à donner à l'utilisateur plus de contrôle sur son fil d'actualité. L'utilisateur du site est invité à connecter son compte Gobo à ses autres comptes de réseaux sociaux (Facebook, Twitter etc.). Gobo récupère ensuite les publications affichées sur ces sites pour proposer un fil d'actualité unifié. La particularité du site repose sur une série de règles de filtrage que l'utilisateur peut librement appliquer à son fil d'actualité. Il peut ainsi décider de masquer certains contenus pour ajuster les proportions de contenus politiques, de contenus « sérieux », de contenus produits par des femmes, de contenus violents, etc. Si le site était en libre accès il y a quelques années, il n'est aujourd'hui plus hébergé. La découverte de ce site est à l'origine du développement d'Alveho. Certaines mécaniques de communication frontend/backend ainsi que la structure en colonne de l'interface utilisateurs sont directement inspirées de ce projet. Gobo a pour limitation principale de ne pas proposer un système de recommandation alternatif. L'ensemble des outils proposés reposent sur un système de catégorisation qui va associer chaque article à une catégorie souvent binaire (sérieux/comique, poli/malpoli, auteur masculin/féminin, républicain/démocrate etc.) et réassortir le fil d'actualité pour que les proportions de chaque typologique correspondent aux demandes de l'utilisateur.

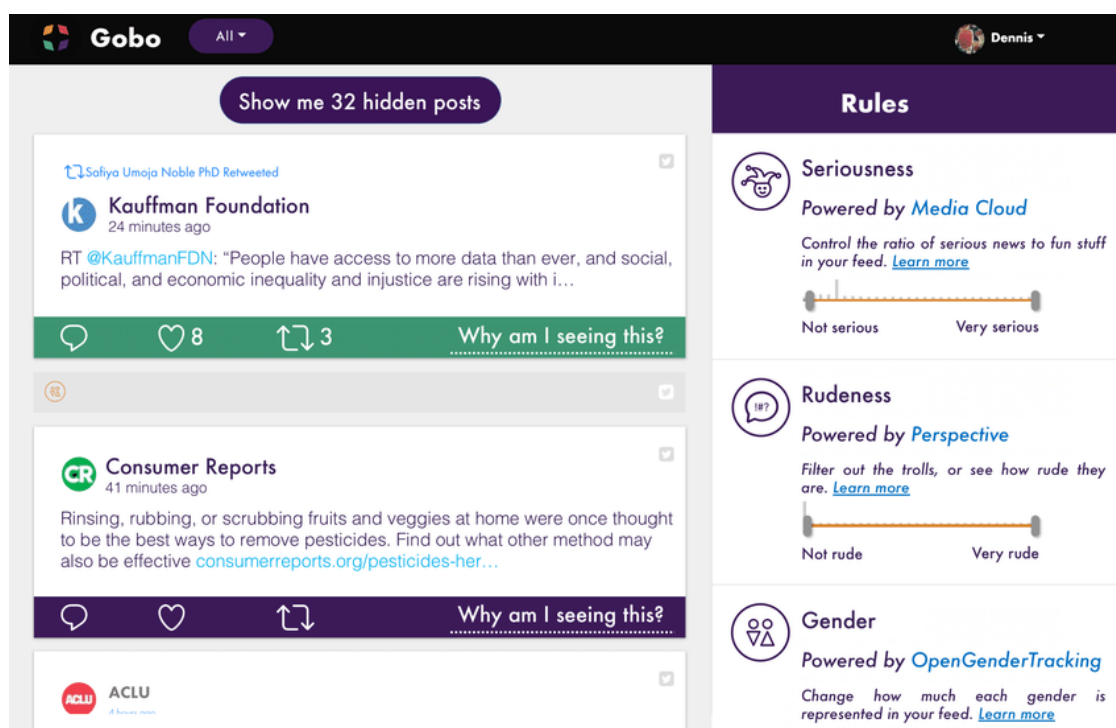


Figure 9 — Gobo

Développée plus récemment, l’extension de navigateur *youchoose.ai* ajoute un nouveau système de recommandation à Youtube. Celui-ci vient s’ajouter à l’interface initiale du site à l’emplacement des contenus recommandés pendant le visionnage d’une vidéo. Les vidéos recommandées par cet outil ne sont pas issues d’un algorithme mais de contenus sélectionnés par le créateur de la vidéo que l’utilisateur est en train de regarder. Les utilisateurs de l’addon peuvent choisir de partager leurs données de navigation (notamment les vidéos qui leur sont recommandées) avec l’équipe de Tracking Exposed, l’équipe de recherche à l’origine du projet. Ce système a pour inconvénient notable de reposer sur la bonne volonté des créateurs de contenus qui doivent volontairement spécifier des vidéos à recommander à leur public. A l’heure où nous écrivons ces lignes, nous n’avons pas été en mesure de trouver de chaînes utilisant ce système à l’exception de celle appartenant à l’équipe de Tracking Exposed.

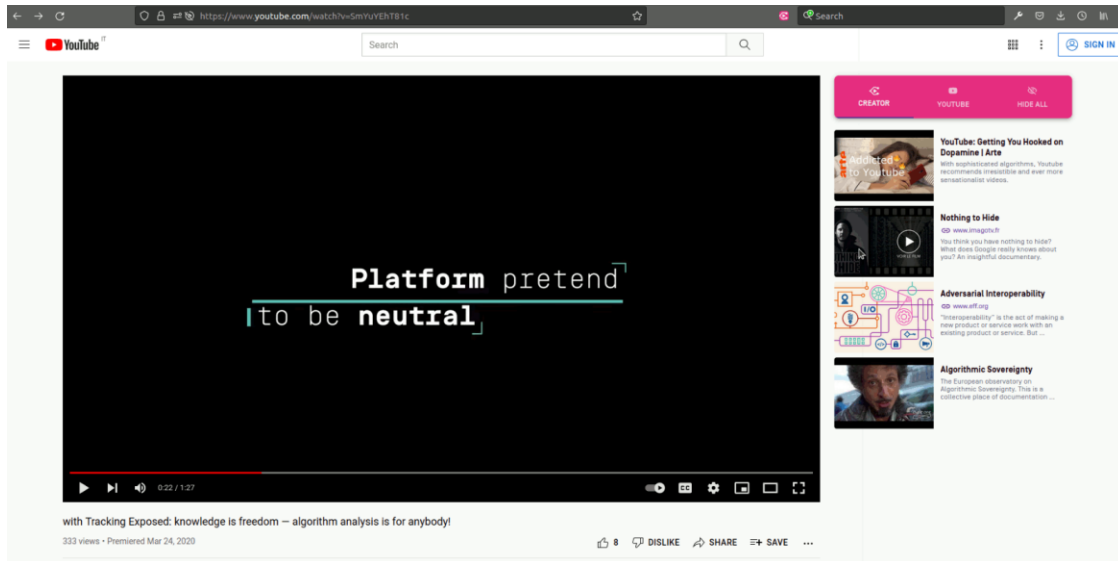


Figure 10 — youchoose.ai

En parallèle de ce premier groupe, d'autres travaux peuvent être associés aux sciences informatiques et plus spécifiquement au domaine de l'interaction homme-machine. Ces initiatives sont généralement peu ou pas médiatisées auprès du grand public et le prototype n'est utilisé que dans le cadre d'études expérimentales. Les résultats de ces travaux sont ensuite publiés dans des revues du domaine. Pour une revue de littérature complète sur ces travaux, nous renvoyons à He *et al.* (2016) ainsi qu'à Jugovac et Jannach (2017). Nous ne présenterons ici que quelques exemples emblématiques.

TasteWeights (Bostandjiev *et al.*, 2012) est un système de recommandation musicale. La colonne de gauche détaille une liste de musiques appréciées par l'utilisateur (les données sont issues de Facebook). La colonne centrale permet à l'utilisateur de sélectionner une série de sources sur lesquelles baser les recommandations. La dernière colonne présente les recommandations classées du haut vers le bas selon leur pertinence. A chaque étape, l'utilisateur peut réajuster les poids de chaque élément pris en compte par l'algorithme à l'aide de curseurs de sorte à changer la liste de recommandations produite.



Figure 11 — Tasteweight issu de Bostandjiev et al. (2012)

Le prototype développé par Jin *et al.* (2018) produit des recommandations musicales. Son interface partage de nombreux points communs avec TasteWeights. La colonne de gauche permet de spécifier des artistes, des pistes musicales ou des genres qui définissent les goûts de l'utilisateur. L'importance de chaque source d'information dans le calcul des recommandations peut être ajustée en modifiant leur poids. Enfin, la colonne de droite présente une série de recommandations que l'utilisateur peut encore filtrer selon ses envies.

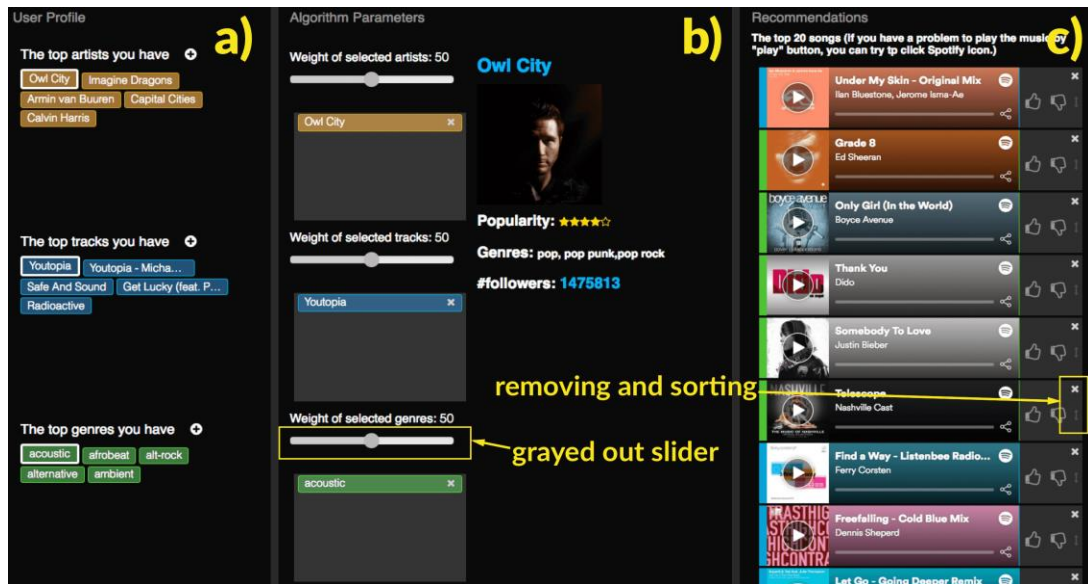


Figure 12 — Prototype développé par Jin et al. (2018)

Setfusion (Parra et Brusilovsky, 2015) est un système de recommandation d'articles scientifiques. Son interface permet à l'utilisateur d'ajuster une série de poids qui déterminent l'importance de différents facteurs impliqués dans le calcul des recommandations. En dessous de ces paramètres, un diagramme de Venn permet de visualiser la proximité de chaque recommandation avec les trois variables que l'utilisateur peut contrôler.

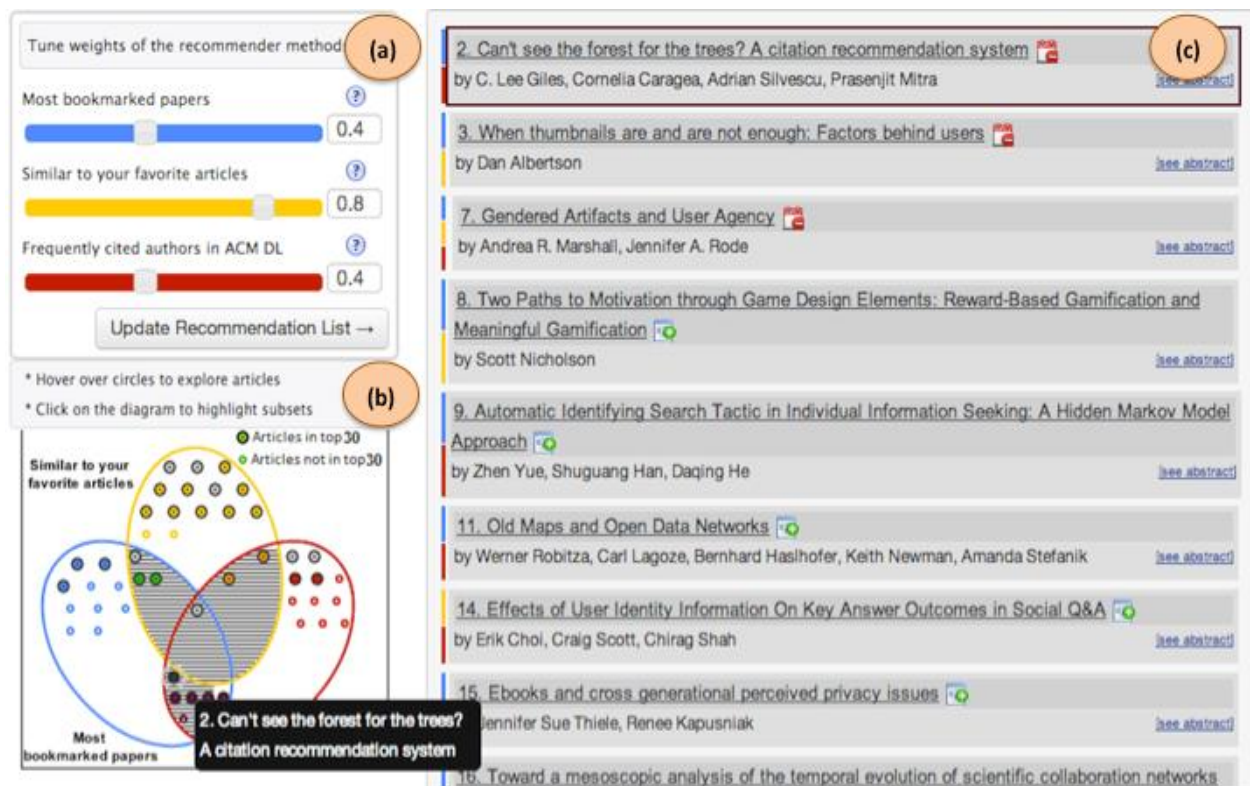


Figure 13 — SetFusion issu de Parra et Brusilovsky (2015)

Ekstrand *et al.* (2015) proposent également un dispositif qui repose sur la plateforme MovieLens, un site de recommandation de films. Les auteurs de l'étude intégrèrent à l'interface de base un système permettant à l'utilisateur de choisir entre plusieurs algorithmes de recommandation, chacun ayant sa propre logique de fonctionnement.

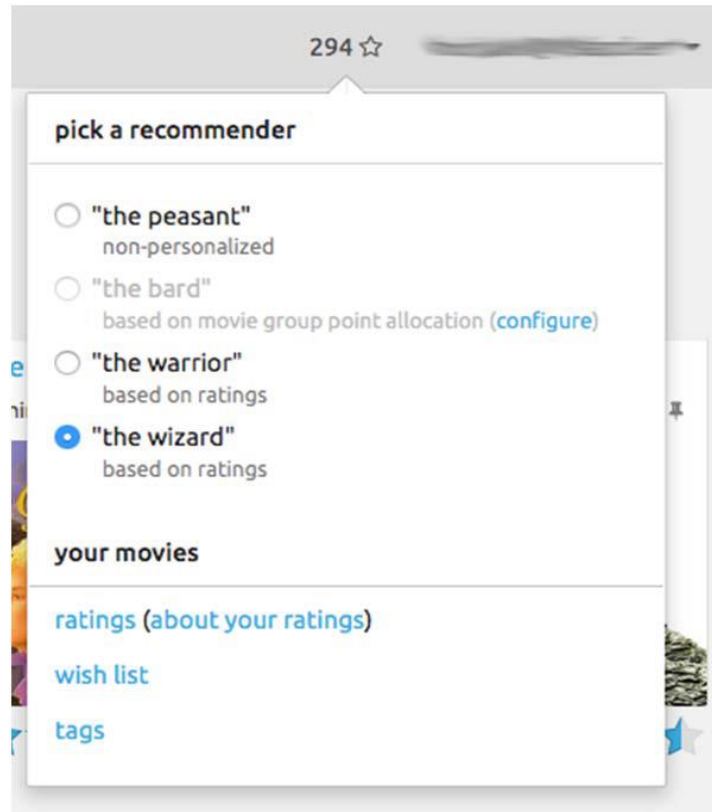


Figure 14 — Choix d'un algorithme sur MovieLens issu de Ekstrand et al. (2015)

Enfin, bien qu'il ne s'agisse pas d'un dispositif fonctionnel mais du mock-up d'une interface, le prototype d'un système de recommandation d'actualités développé par Harambam *et al.* (2019) partage de nombreuses similitudes avec notre prototype. L'interface proposée permet de spécifier ses préférences, de choisir l'algorithme de recommandation et de filtrer les recommandations une fois que celles-ci ont été calculées par l'algorithme.



Figure 15 — Mock-up développé par Harambam et al. (2019)

En Belgique, la RTBF travaille également à l'élaboration d'outils de contrôle d'un algorithme de recommandation. Les captures d'écran ci-dessous présentent un prototype interne reposant sur la sélection d'une stratégie de recommandation parmi une liste de plusieurs algorithmes.

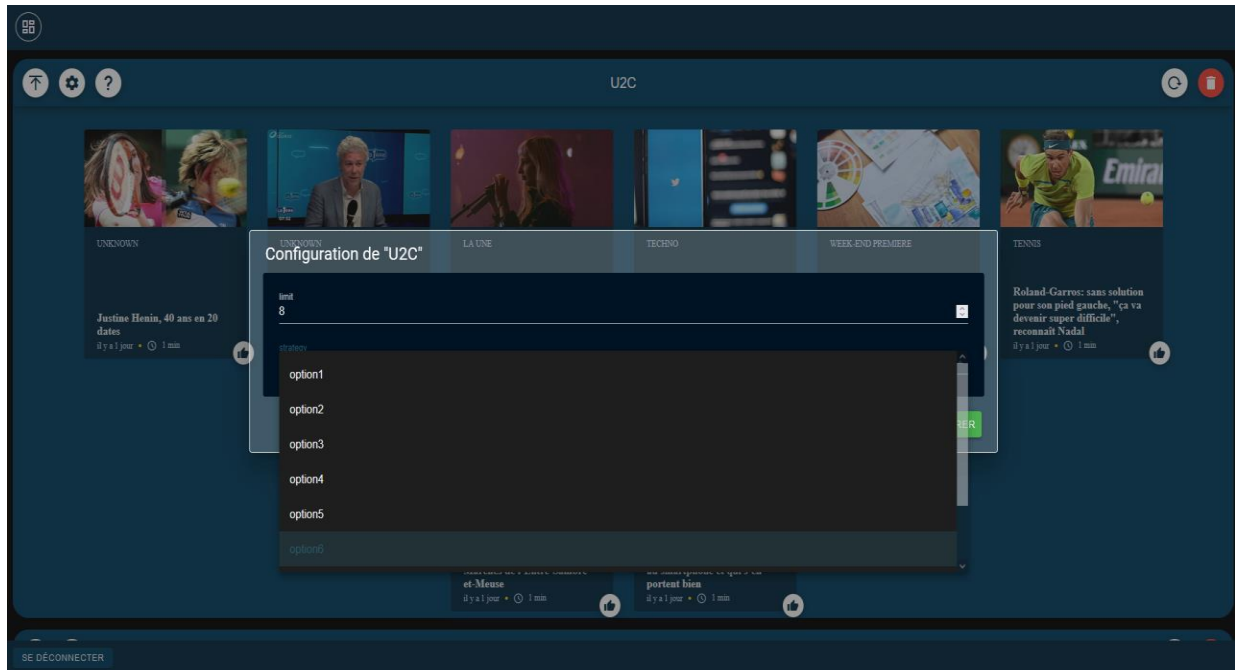


Figure 16 — Capture d'écran d'un prototype interne à la RTBF - Options de contrôle

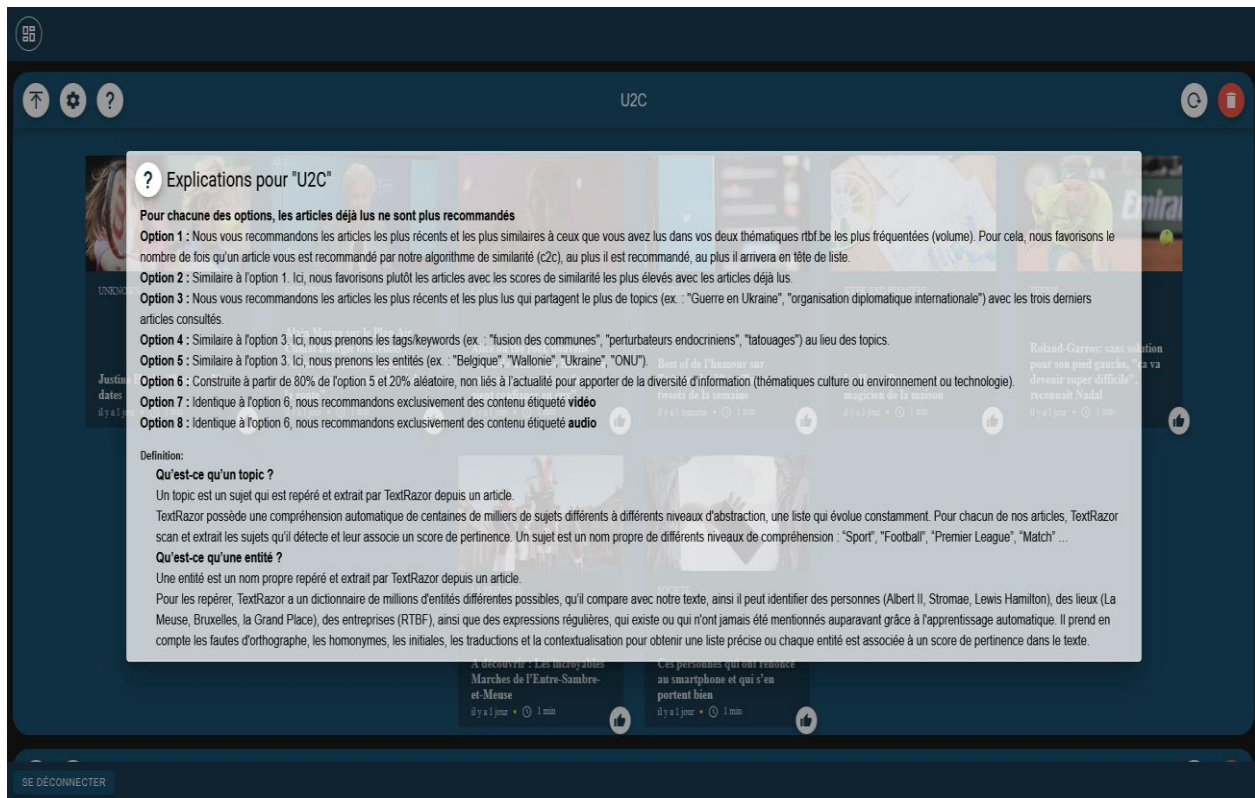


Figure 17— Capture d'écran d'un prototype interne à la RTBF - Explication des options de contrôle

Étant donné leur ancrage disciplinaire, les recherches présentées dans cette section sont principalement préoccupées par l'expérience utilisateur, c'est-à-dire « la façon dont un individu va pouvoir appréhender les différentes fonctions techniques et va ensuite mettre en œuvre les scripts d'action offerts par la technologie en fonction de buts et d'objectifs préétablis » (Denouël, 2012, p. 74). Cette expérience utilisateur est habituellement décomposée en 3 dimensions principales : utilité (adéquation entre les fonctionnalités et les besoins de l'utilisateur), utilisabilité (efficacité et satisfaction d'utilisation) et acceptabilité (correspondance entre les caractéristiques du système et les conventions sociales des situations d'utilisation) (Tricot *et al.*, 2003). Les recherches précitées s'intéressent spécifiquement à l'utilisabilité à travers la mesure de la satisfaction de l'utilisateur et la précision du système. Leurs résultats semblent indiquer un effet positif de la contrôlabilité de l'algorithme sur ces deux variables. A l'exception des travaux d'Harambam *et al.* (2018), la littérature issue de ces recherches ne se croise pas avec les travaux du domaine portant sur la réalisation de soi présentés dans le chapitre 2 (Knijnenburg *et al.*, 2016).

Dans la majorité des cas, il est explicitement demandé aux participants de l'étude d'interagir avec les fonctionnalités du prototype. A l'exception des travaux d'Ekstrand *et al.* (2015), nous ne disposons que de peu d'information sur le taux d'utilisation de ces fonctionnalités dans un contexte d'usage plus libre. En ce qui concerne cette étude, sur les 3005 participants, 748 (24,9 %) d'entre eux ont changé d'algorithme au moins une fois. 539 participants (72,1 % des utilisateurs ayant utilisé l'outil de contrôle) se sont arrêtés sur un algorithme différent que celui qui leur fut assigné au départ. Les auteurs de l'étude en concluent que les utilisateurs d'un système de recommandation influent sur le système quand l'opportunité leur en est donnée. Sachant que 75,1 % des participants n'ont pas utilisé l'outil de contrôle, il nous semble que ces résultats prouvent l'hypothèse inverse.

Typologie des modalités d'interaction avec le système

Nous proposons de reprendre la typologie utilisée par Harambam *et al.* (2019) qui divise les fonctionnalités de contrôle d'un algorithme de recommandation en trois groupes en fonction de leur moment d'intervention dans le processus de recommandation :

- 1) Lors de la spécification des inputs (l'explicitation des préférences de l'utilisateur) ;
- 2) Lors du processus de calcul de l'algorithme ;
- 3) Lors de l'émission des outputs (la présentation des recommandations).

Ces trois types d'interactions avec l'algorithme sont généralement accompagnés par des explications. Si celles-ci peuvent opérer en parfaite indépendance avec les outils de contrôle, elles font parfois partie intégrante de la manipulation du système.

Input

Il est assez rare de demander à l'utilisateur d'explicitement ses préférences avant d'utiliser un système de recommandation. La plupart des systèmes vont plutôt se baser sur des signaux implicites pour émettre leurs recommandations (Ekstrand et Willemsen, 2016). La récolte de données est généralement invisible pour l'utilisateur ce qui contribue à masquer la présence d'un algorithme et à opacifier son fonctionnement (Eslami *et al.*, 2015).

Lorsque de telles fonctionnalités sont implémentées, elles apparaissent principalement à la première connexion pour pallier aux problèmes de démarrage à froid (« cold start » dans le jargon des ingénieurs). Toutefois, quelques travaux se sont intéressés à l'implémentation d'outils de contrôle au niveau des inputs pour permettre à l'utilisateur de mieux comprendre ses comportements et d'explicitement ses préférences (Graus *et al.*, 2018). Sullivan *et al.* (2019) suggèrent que de telles fonctionnalités pourraient soutenir des pratiques d'actualisation de soi mais ne fournissent pas de données empiriques sur le sujet. Toutefois, les résultats de Karakayali *et al.* (2018) semblent indiquer que permettre à l'utilisateur de sélectionner de façon consciente les données qu'il fournit au système contribue à les conscientiser sur leurs goûts et leurs intérêts.

Processus

La majorité des travaux relatifs à l'implémentation de paramètres de contrôle se focalisent sur le contrôle direct de l'algorithme. Deux stratégies générales peuvent être identifiées : la modification de poids ou la sélection d'une logique générale de recommandation.

La première stratégie consiste à structurer l'algorithme de recommandation autour d'une formule où chaque paramètre est associé à un poids. Ce poids est ensuite rendu ajustable au niveau de l'interface utilisateur. Cette méthode a été adoptée par Bostandjiev *et al.* (2012), Harper *et al.* (2015), Jin *et al.* (2018), Knijnenburg, Bostandjiev, *et al.* (2012) ainsi que Parra et Brusilovsky (2015). Une autre stratégie consiste à proposer à l'utilisateur de choisir entre plusieurs algorithmes différents et lui permettre de changer ce choix à la volée selon ses préférences. (Ekstrand *et al.*, 2015). Cette méthode ne permet pas à l'utilisateur d'avoir un contrôle sur la recommandation aussi

fin qu'avec des poids. Toutefois, elle permet d'expérimenter des logiques de recommandation moins conventionnelles (Wilkinson, 2018). C'est notamment l'option explorée par l'extension de navigateur youchoose.ai. L'utilisateur ne peut pas vraiment configurer l'algorithme mais il peut alterner entre deux logiques différentes (sélection algorithmique ou sélection humaine).

Plusieurs recherches ont constaté que fournir un contrôle direct sur l'algorithme influe positivement sur la précision des recommandations ainsi que sur la satisfaction de l'utilisateur (Bostandjiev *et al.*, 2012 ; Harper *et al.*, 2015). Toutefois, tous les utilisateurs n'ont pas recours à ces fonctionnalités quand elles sont disponibles (Ekstrand *et al.*, 2015 ; Harper *et al.*, 2015) et les facteurs qui pourraient expliquer cette disparité ne sont pas encore identifiés.

Output

Une fois que les recommandations sont produites, il est encore possible de réorganiser les résultats en fonction de l'une ou l'autre préoccupation. L'approche la plus commune consiste à supprimer des objets de la liste (Bostandjiev *et al.*, 2012 ; Jin *et al.*, 2018). Les fonctionnalités proposées par Gobo interviennent principalement au niveau de l'output. Le système ne contrôle pas un algorithme de recommandation mais propose plutôt de filtrer des contenus selon une série de priorités sélectionnées par l'utilisateur.

Explications

L'implémentation de fonctionnalités de contrôle s'accompagne généralement d'outils explicatifs permettant de contextualiser les suggestions. Celles-ci peuvent soit prendre la forme d'explications textuelles, de visualisations de données ou de statistiques (Herlocker *et al.*, 2000). Dans certains cas, il est possible d'interagir avec ces visualisations pour modifier le comportement de l'algorithme. Parra et Brusilovsky (2015) ont ainsi recours à un diagramme de Venn pour permettre à l'utilisateur de filtrer certains contenus recommandés. Ces fonctionnalités sont généralement évaluées pour leur influence sur la satisfaction des utilisateurs et la qualité perçue des recommandations (Knijnenburg, Willemsen, *et al.*, 2012). Nous n'avons pas été en mesure d'identifier de recherches qui s'intéressent à l'effet de ces explications sur le développement de compétences réflexives de la part des utilisateurs.

Alveho²³ — Interface utilisateur

En nous inspirant des travaux cités précédemment, nous avons développé notre propre plateforme expérimentale, intitulée Alveho et incorporant son propre système de recommandation. Les pages qui suivent sont dédiées à la présentation des fonctionnalités principales de cette plateforme. Des captures d'écran complètent nos explications. Une vidéo d'Alveho ainsi que des captures d'écran supplémentaires sont disponibles dans les annexes 5 et 6.

Algorithme de recommandation

Alveho affiche un fil d'actualité composé d'articles de presse issus d'un média de service public belge francophone (RTBF). Ce fil d'actualité est affiché au centre de l'écran à l'image de réseaux sociaux classiques tandis que les autres fonctionnalités du site se disposent de part et d'autre de cette colonne centrale.

Les articles de ce fil d'actualité sont classés par un algorithme de recommandation développé avec l'aide du Centre du traitement automatique du langage (CENTAL) de l'UCLouvain. Lorsqu'un utilisateur se connecte, 5000 articles sont récupérés dans la base de données²⁴ et se voient attribués un score de pertinence. Plus la valeur sera élevée, plus l'article sera classé haut dans le fil d'actualité. Cette valeur dépend de deux variables principales qui sont additionnées entre elles et multipliées par des poids déterminés par l'utilisateur.

Ce score est calculé de la façon suivante :

$$r_{u,i} = w_u^0 \cdot sim_{u,i} + w_u^1 \cdot rec_i$$

²³ Le développement de la version d'Alveho présentée dans ce chapitre prit 8 mois (approximativement du 15 février au 15 octobre 2022) et mobilisa 3 développeurs. Les tâches de développement furent réparties de la façon suivante :

1. Claes Arnaud (UCLouvain, GReMS, Chercheur de l'étude) : Design de la plateforme, coordination générale, backend (base de données, API, sécurité, algorithme), frontend (sécurité, outils de contrôle, visualisations de données), prétest de la plateforme ;
2. Maxime Janssens (UCLouvain, Social Media Lab, Développeur web junior) : backend (Mise en place de flask, gestion utilisateurs), Frontend (Mise en place vue.js, homepage, fil d'actualité) ;
3. Damien De Meyere (UCLouvain, Cental, Développeur en traitement automatique du langage) : Outils de traitement automatique du langage (analyse de similarité, mots-clés, recherche), mise en place du serveur.

²⁴ Pour la justification quant à cette valeur, voir la sous-section sur l'architecture interne d'Alveho

Où :

- $r_{u,i}$ est le score de l'objet i pour l'utilisateur u
- w_u^0 est le poids du score de similarité
- $sim_{u,i}$ est la similarité entre un objet i et l'historique de l'utilisateur u
- w_u^1 est le poids du score de récence
- rec_i est le score de récence pour un objet i

La similarité entre deux articles est obtenue en calculant la distance cosine entre une représentation vectorielle de chaque article. Cette représentation vectorielle est obtenue en utilisant une allocation de Dirichlet latente qui génère une distribution de probabilité des sujets de chaque article. Le score de récence est donné par une fonction dont le résultat décroît avec l'âge de l'article. L'utilisateur peut attribuer aux poids une valeur allant de -10 à 10. Une valeur négative implique que l'utilisateur renverse la priorité du système. Par exemple, une similarité négative va donc favoriser des articles qui diffèrent des habitudes de consommation de l'utilisateur.

Ce mode opératoire permet de faciliter l'intégration future de paramètres additionnels qui doivent simplement être ajoutés à la suite de la formule existante.

Page d'accueil

Une fois que l'utilisateur est connecté à son compte personnel, il est automatiquement redirigé vers la page principale du site (voir figure 18). Celle-ci se décompose en trois colonnes surplombées par une barre de navigation. Celle-ci comprend un menu, un raccourci vers la page de profil et un autre raccourci vers un manuel expliquant les fonctionnalités principales du site.

La colonne centrale est un fil d'actualité (2) composé d'articles de presse produits par la RTBF. Chaque article dispose d'une catégorie. Au moment de la réalisation de cette étude, ces catégories correspondaient soit à une typologie de contenu issue de la « chaîne principale » de la RTBF (RTBF Info, RTBF Tendances, RTBF Culture, RTBF Sport, RTBF radio, RTBF TV) soit à une sous-chaîne appartenant à l'organisme (Tipik, Classic21, LaPremière, Vivacité). Cette situation conduit à une typologie confuse qui mélange des typologies générales de contenu avec des chaînes d'information à l'identité éditoriale marquée (Classic21 orientée musique rock, Tipik orientée vers un public

jeune, etc.). Suite à un changement dans l'architecture de la RTBF qui a eu lieu après notre expérimentation, la typologie fut remaniée et dispose d'une cohérence interne plus compréhensible. Chaque article dispose également d'un deuxième niveau de catégorisation plus fine. Néanmoins, ces catégories sont ajoutées manuellement par des éditeurs de contenu de la RTBF ce qui conduit à la prolifération de plusieurs centaines de mots-clés, parfois reflétant le même objet mais orthographié différemment. L'absence de cohérence dans ces catégories de second niveau les rend inutilisables dans le cadre de cette étude.

Lorsque l'utilisateur clique sur la vignette d'un article, une fenêtre s'ouvre pour lui permettre de consulter l'article. À l'exception de quelques contenus issus de réseaux sociaux qui ne s'affichaient pas correctement, les articles présentés sur Alveho comprenaient l'ensemble des contenus multimédias visibles sur le site de la RTBF (images, vidéos, etc.). Ceci garantit un certain confort de lecture proche de l'expérience réelle sur le site du média de service public.

À gauche de ce fil d'actualité (1), on peut retrouver un graphique circulaire qui montre la répartition des articles du fil d'actualité selon leur catégorie principale. En dessous de ce graphique, nous avons ajouté un nuage de multi-mots qui, contrairement à un nuage de mots classiques, va extraire des groupes de mots conjoints depuis les articles présents dans le fil d'actualité. Conformément à un nuage de mots classiques, plus le mot ou le groupe de mots sera grand, plus son nombre d'occurrences dans le fil d'actualité sera important. Ces deux outils visent à fournir une vue d'ensemble à l'utilisateur sur les contenus présents dans la colonne recommandée.

À droite du fil d'actualité, les deux outils les plus notables sont les paramètres de contrôle de l'algorithme (5) ainsi que l'outil de filtrage du flux par catégorie. Le premier permet d'ajuster les deux réglages principaux de l'algorithme (similarité et actualité) tandis que le second permet de cacher ou d'afficher certains articles selon leur catégorie (6). Si l'ajustement des paramètres de l'algorithme implique un rafraîchissement du fil d'actualité et la demande d'une nouvelle liste d'articles, ce n'est pas le cas pour le filtrage par catégorie qui ne fait que masquer certains articles issus de la dernière requête effectuée.

Dans la même colonne, se trouve également un outil de recherche (3). Lorsque l'utilisateur commence à taper un mot, un outil d'auto-complétion lui propose une série de propositions sur base des (multi-)mots clés extraits des articles. La sélection d'un de ces mots est obligatoire. Il n'est donc pas possible de soumettre un mot qui n'est pas dans la liste de mots-clés.

Enfin, en dessous de l'outil de recherche, un bouton permet de remplacer le fil d'actualité par une copie de la colonne de gauche, à la fois le graphique circulaire et le nuage de mots (4). Cette copie diverge de l'original en utilisant une source de données différente. Au lieu de représenter le contenu du fil d'actualité, cette copie va décrire le contenu d'un hypothétique fil d'actualité organisé par ordre antéchronologique.

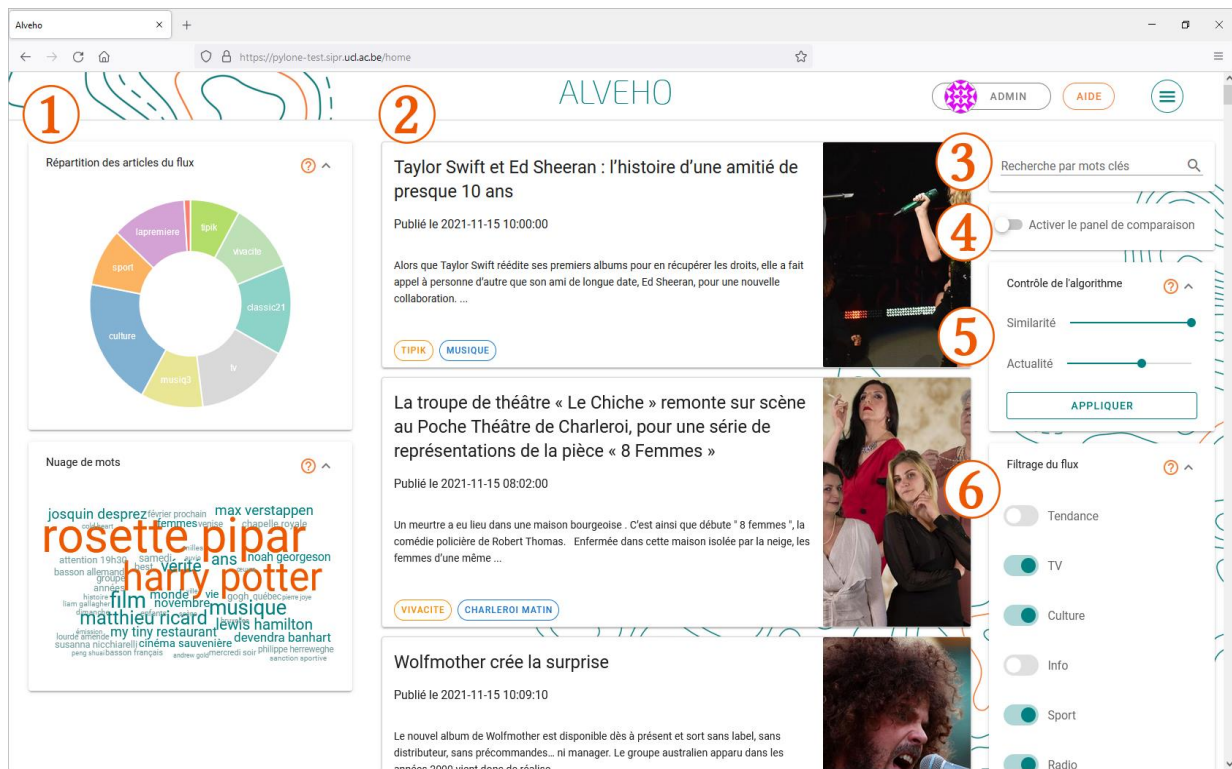


Figure 18 — La page principale d'Alveho



Figure 19 — Fenêtre de lecture d'un article sur Alveho

Page de profil

La page de profil (voir figure 20), accessible via les raccourcis de la barre de navigation, permet à l'utilisateur de consulter et de manipuler son fil d'actualité. Parmi les quatre composants principaux, deux sont des visualisations de données et deux autres sont des outils de filtrage de l'historique. Lorsqu'un article enregistré dans l'historique de l'utilisateur est filtré par l'un de ces composants, il disparaît de la liste prise en compte par l'algorithme pour réaliser les recommandations. Admettons que nous avons un utilisateur qui lise à la fois des articles culturels et des articles sportifs, s'il décide de filtrer les articles sportifs, il ne devrait recevoir que des recommandations correspondant à un registre culturel. Les deux graphiques s'actualisent en temps réel à chaque fois qu'un paramètre change de sorte à représenter l'état actuel de l'historique de lecture.

Le graphique en zones empilées (1) permet de visualiser l'évolution de l'historique de navigation en fonction du temps. Chaque zone colorée correspond à une catégorie. L'axe vertical s'exprime en pourcentage tandis que chaque graduation de l'axe horizontal représente un jour. La ligne horizontale située en dessous du graphique permet de filtrer l'historique de navigation en fixant une date minimale et une date maximale (3). Tous les articles lus en dehors de ces deux bornes sont

cachés à l'algorithme. Le graphique circulaire (2) à droite de l'écran représente la répartition par catégorie des articles inscrits dans l'historique de navigation. Il représente donc la même chose que la dernière tranche du graphique en zones empilées. Les boutons situés en dessous de ce graphique permettent de filtrer l'historique de lecture selon les catégories des articles (4).

Enfin, nous avons comblé l'espace restant dans l'interface par un résumé de l'activité quotidienne de l'utilisateur (5). Il s'agit d'un texte à trous dans lequel certaines statistiques de base sont insérées comme le nombre d'articles publiés, le nombre d'articles moyens consultés par l'utilisateur ou encore la moyenne des autres utilisateurs du site. C'est également à cet endroit que nous avons ajouté un indicateur du temps passé sur le site. Plutôt que d'indiquer le temps exact, l'indicateur passe simplement au vert lorsqu'un seuil minimal est atteint.

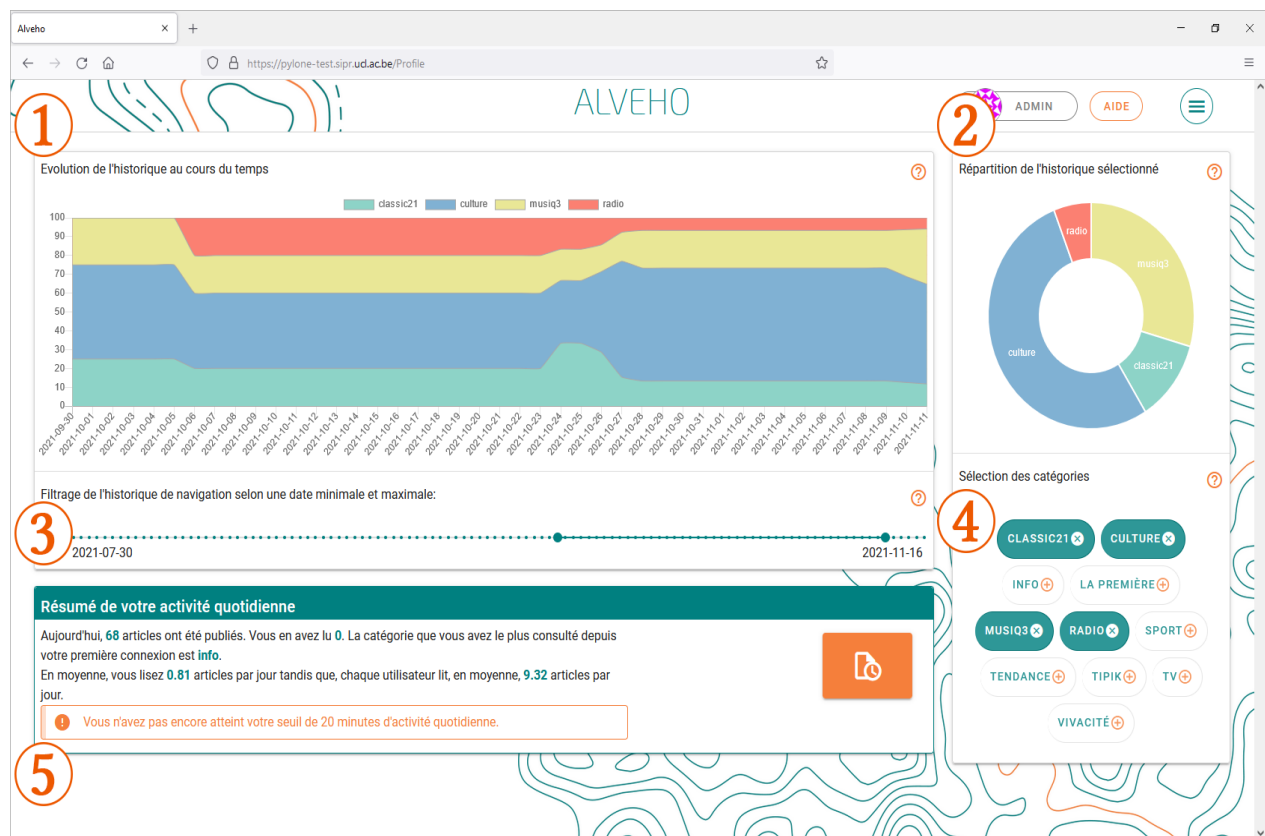


Figure 20 — La page de profil d'Alveho

Fonctionnalités d'administration

Pour faciliter la gestion du site pendant une expérimentation, nous avons implémenté un premier prototype d'interface de gestion dont l'accès est réservé à un compte administrateur. Cette interface permet de réaliser trois types d'opérations majeures:

- 1) Créer un nouveau compte utilisateur sur base d'une adresse email.
- 2) Supprimer, désactiver ou réactiver un compte utilisateur. Un compte utilisateur désactivé existe toujours dans la base de données, mais l'accès au site est refusé. Son token n'est pas pour autant invalidé.
- 3) Extraire sous format CSV les logs enregistrés dans la base de données. À l'heure actuelle, il est uniquement possible d'extraire l'ensemble des logs en une seule fois ou de n'extraire que les logs de suivi de l'activité.

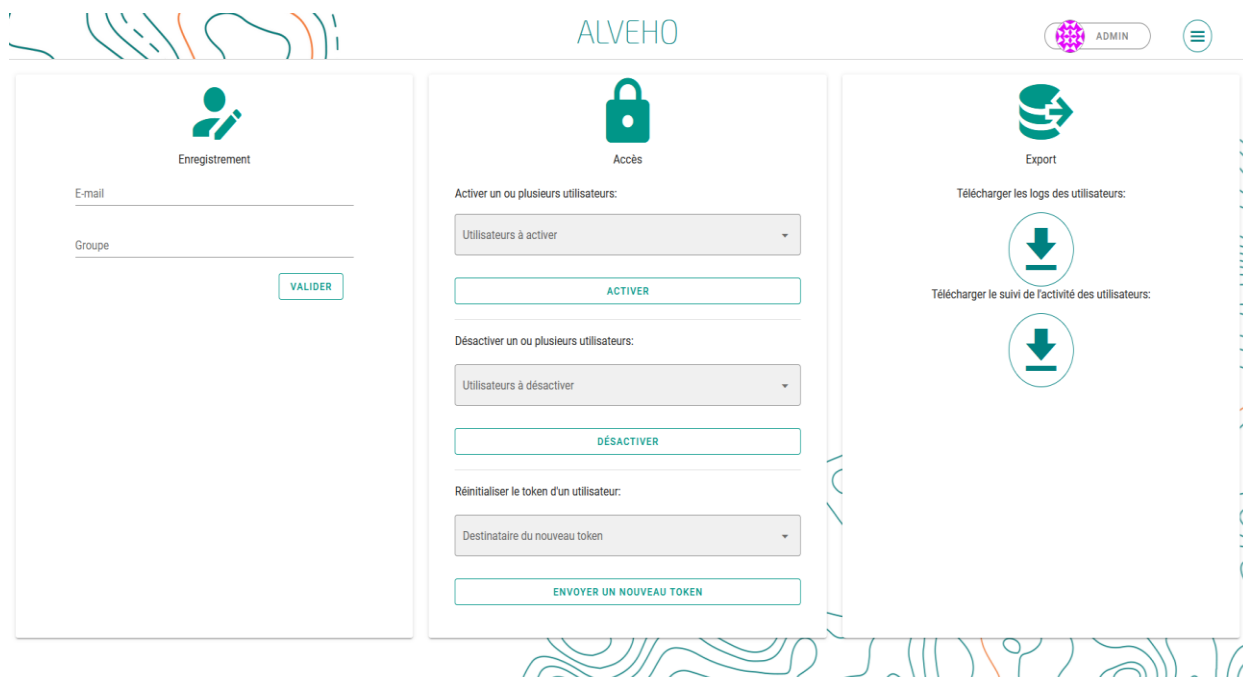


Figure 21 — Page d'administration d'Alveho

Alveho — Architecture interne

Il nous semblait important de revenir sur les mécanismes internes de la plateforme étant donné l'enchevêtrement complexe entre notre méthodologie et l'architecture d'Alveho. Si certaines

fonctionnalités découlent d'une intention première imposée par notre cadrage méthodologique, certaines contraintes techniques sont venues en retour peser sur nos objectifs de recherche.

Étant donné notre volonté de placer les participants de l'étude dans des conditions les plus proches possibles de conditions réelles, nous avons dû porter une attention particulière à l'expérience utilisateur afin de masquer autant que possible la nature expérimentale du dispositif. Plusieurs compromis furent effectués entre les fonctionnalités souhaitées et ce qu'il nous était possible de réaliser compte tenu de nos ressources.

Les technologies suivantes furent employées pour développer Alveho :

- La librairie JavaScript vue.js pour l'interface ;
- La librairie python Flask pour la gestion du serveur ;
- Postgresql pour la gestion de la base de données.

Récupération des articles

Les articles qui alimentent la base de données sont directement récupérés depuis la base de données de la RTBF à travers leur API. En tâche de fond de l'application principale qui gère Alveho, un planificateur de tâches envoie une requête aux serveurs de la RTBF toutes les 15 minutes pour récupérer les articles publiés depuis la dernière requête. Le catalogue de la RTBF est séparé en différents flux en fonction de la chaîne ou de la catégorie à laquelle appartient l'article. Nos requêtes récupèrent les articles publiés sur l'ensemble de ces flux.

Lors de chaque requête sont récupérés l'identifiant de l'article, le contenu de l'article, la date de publication, l'URL de l'image d'en-tête, le nom du flux, la catégorie associée quand il y en a une et l'auteur de l'article. Récupérer l'identifiant original de l'article permet de vérifier que l'article n'a pas déjà été récupéré lors d'une autre requête. Chaque article est décomposé par la RTBF en une liste de blocs. Chaque bloc correspond à un média composant l'article (une image, un paragraphe, une vidéo, un message issu d'un réseau social, etc.). Cette liste de blocs est stockée au format JSON²⁵ tel que nous l'obtenons.

²⁵ JSON pour JavaScript Object Notation est un format de données textuelles qui permet de représenter des informations de façon structurée à l'image du format XML.

Le texte de l'article est recomposé sans les autres médias présents dans la liste de blocs. Ce texte brut sert au calcul du vecteur de probabilités à la base de la mesure de similarités entre articles ainsi qu'à l'extraction des mots-clés alimentant le nuage de mots et le module de recherche. Ce vecteur ainsi que le texte brut sont stockés dans la base de données en même temps que les autres attributs récupérés depuis les serveurs de la RTBF.

Gestion des utilisateurs

Afin de faciliter l'accès à la plateforme aux participants de l'étude et afin de limiter les étapes intermédiaires, nous avons opté pour une méthode d'authentification par token d'identification plutôt que par mot de passe. La création de comptes n'est pas ouverte aux utilisateurs et ne peut être réalisée que par l'administrateur du site en encodant l'adresse email d'un participant. Lorsqu'un compte est créé, un JSON Web Token d'identification unique est généré sur base de cette adresse, de la date de création du compte et d'une clé secrète stockée sur le serveur.

Chaque jeton d'identification comprend un en-tête décrivant le jeton, une charge utile qui représente les informations stockées dans le jeton (notamment l'adresse email) et une signature numérique. Concrètement, l'en-tête et la charge utile sont encodés en Base64. Les chaînes de caractères issus de l'encodage sont concaténées et séparées par un point. Cette chaîne concaténée est ensuite cryptée sur base de la clé secrète du serveur. Le résultat est également encodé en Base64 avant d'être concaténé à l'en-tête et à la charge utile. Ceci implique que n'importe qui disposant du jeton peut connaître le contenu du token mais toute modification invaliderait la signature numérique qui ne correspondrait plus à la combinaison de la charge utile et de l'entête.

Chaque participant reçoit un mail comprenant l'URL du site à laquelle est accolé le token généré. Lorsque l'utilisateur accède au site depuis cette adresse, le token est récupéré et envoyé au serveur afin que l'identité de l'utilisateur soit vérifiée. La signature numérique est décodée à l'aide de la clé secrète. Si le décodage du token d'identification génère une adresse email stockée dans la base de données, l'utilisateur est connecté au site au compte associé à l'adresse email. Lors de la première connexion, un cookie est déposé sur le navigateur du participant pour le connecter immédiatement lors de ses prochaines connexions au site. Si ce cookie est supprimé, il suffit au participant de repasser par le lien. En supposant que le participant est attentif à ne pas partager ce jeton, la sécurité de cette méthode est dépendante de la sécurité de la boîte mail de l'utilisateur. De plus, avec cette méthode, aucune information d'authentification n'est stockée sur le serveur.

Étant donné les contraintes de l'étude, une copie de chaque jeton fut diffusée sur la console du serveur lors de la création des comptes. Avant le nettoyage de la console, les jetons furent récupérés et stockés sur un disque dur sécurisé. Disposer d'une copie des jetons nous permet de nous connecter aux comptes des participants, ce qui a deux avantages majeurs :

- 1) Plusieurs bugs inattendus sont apparus lors de la première semaine d'expérimentation. Accéder directement au compte des participants nous a permis de corriger ces erreurs plus rapidement.
- 2) Lors des entretiens réalisés pour l'étude qui a suivi le développement, disposer de ces clés nous a permis de connecter les comptes des participants sur l'ordinateur utilisé pour l'occasion.

En fonction de la portée et des enjeux d'études futures, il peut être préférable de ne pas disposer d'une copie des jetons d'authentification.

Gestion des logs

Lorsque l'utilisateur interagit avec certaines fonctionnalités du site, un signal est envoyé au serveur pour enregistrer un log de l'action réalisée. Les logs sont enregistrés au format suivant : horodatage, identifiant de l'utilisateur, action réalisée, valeur. Un log est généré chaque fois que l'utilisateur ajuste le filtrage des catégories, le filtrage de l'historique, les paramètres de l'algorithme mais aussi quand il réalise une recherche ou qu'il ouvre le manuel.

Mesure de la durée de l'activité en ligne

Pour assurer que le taux de participation reste régulier pendant l'expérimentation, nous avons décidé d'intégrer un système qui mesure, pour chaque utilisateur, le temps passé à naviguer sur le site. Il est possible de capter un signal lorsque l'utilisateur se connecte au site. Par contre, il n'est pas possible de savoir avec certitude quand l'utilisateur ferme la page du site étant donné qu'il coupe toute forme de communication avec le serveur. De plus, un utilisateur peut garder sa fenêtre de navigation ouverte sur le côté tout en réalisant une autre activité.

Pour ces raisons, nous avons implémenté un système de suivi de l'activité de l'utilisateur. Si après 60 secondes, le navigateur client n'observe plus d'activité de la part de l'utilisateur (bouger la souris, faire défiler la page, appuyer sur une touche, toucher l'écran dans le cas d'un écran tactile), un signal est envoyé au serveur indiquant l'inactivité de l'utilisateur. Un nouveau signal est envoyé

au serveur dès que l'utilisateur réalise une nouvelle action après avoir été détecté comme inactif. Dans les deux cas, un log est enregistré pour conserver l'information. Lors du calcul du temps d'activité totale, si le temps écoulé entre deux logs déclarant une activité de l'utilisateur est supérieur à 120 secondes, l'utilisateur est considéré inactif pour cet intervalle. Cette implémentation est rendue possible par le faible nombre d'utilisateurs, lui-même lié à la nature expérimentale du site, qui permet de garder une charge de travail légère pour le serveur.

Si ce système est suffisamment précis dans la majorité des cas, il perd en efficacité lorsque l'utilisateur connecté ne dispose pas d'une connexion internet fiable. Dans ces cas-là, certains signaux ne sont pas envoyés au serveur et le temps de navigation calculé est inférieur au temps réel passé sur le site. Cela a conduit certains utilisateurs à passer plus de temps que nécessaire sur le site. Par contre, l'inverse n'est jamais arrivé. Ce point est plus longuement discuté dans la section de discussion de la méthode du chapitre 4.

Démarrage à froid

Un problème récurrent lors de l'usage d'un système de recommandation concerne les recommandations qui sont effectuées lors des premières interactions de l'utilisateur avec le système. L'expression « démarrage à froid » désigne la période pendant laquelle l'algorithme ne dispose pas d'assez d'informations pour proposer des recommandations à l'utilisateur. Dans notre cas, tant que l'utilisateur n'a pas consulté d'articles et que son historique de lecture est vide, le système ne peut pas calculer de mesure de similarité entre son profil et un article de la base de données.

Une solution classique consiste à demander à l'utilisateur de spécifier ses préférences lors de sa première connexion afin d'alimenter le système avec suffisamment de données sur l'utilisateur pour réaliser une première série de recommandations. Nous avons donc implémenté un premier prototype d'élicitation de préférences qui demande à l'utilisateur de sélectionner une première série d'articles qui sont par la suite ajoutés à son historique de navigation. Lors de sa première connexion, l'utilisateur se voit proposer une liste de titres d'articles correspondant à des articles de la base de données. Ces articles sont répartis par catégorie et chaque catégorie est dotée de 3 articles. Il est ensuite demandé à l'utilisateur de sélectionner les catégories qui l'intéressent le plus sur base des titres proposés. Les 3 articles associés aux catégories sélectionnées sont ensuite ajoutés à son historique de navigation.

Toutefois, les tests utilisateurs préliminaires nous indiquèrent plusieurs problèmes d'utilisabilité relatifs à notre implémentation. Les participants de l'évaluation ne comprenaient pas pourquoi des articles qu'ils n'avaient pas réellement lus se retrouvaient dans leur historique de navigation. Étant donné nos contraintes temporelles, il fut décidé de supprimer cette fonctionnalité et de proposer une série de recommandations aléatoires à l'utilisateur lors de sa première connexion. L'expérimentation principale a ensuite révélé une certaine confusion provoquée par l'aléa des recommandations lors de la première connexion. Après l'expérimentation, ce système fut finalement adapté pour proposer les derniers articles publiés par ordre antéchronologique.

Gestion de l'affichage

Au chargement de la page d'accueil, 100 articles sont demandés au serveur et stockés sur le navigateur. Ces articles sont sélectionnés par notre algorithme de recommandation au sein des 5000 derniers articles ajoutés à la base de données. Ces valeurs furent fixées pour des raisons de performances afin que l'utilisateur ne soit pas confronté à des temps de latence trop importants. Chaque paramètre que sélectionne l'utilisateur est enregistré dans la base de données afin de pouvoir rétablir, à chaque nouvelle connexion, l'interface dans l'état laissé par l'utilisateur à sa dernière connexion.

Lorsque l'utilisateur ajuste des filtres de la page d'accueil, les articles présentés sont modifiés à la volée sans émettre de nouvelles requêtes à la base de données afin de ne pas saccader l'expérience utilisateur. Par contre, l'ajustement des paramètres de l'algorithme nécessite l'envoi d'une requête au serveur afin de récupérer une nouvelle série de recommandations correspondant aux nouveaux paramètres demandés par l'utilisateur. Ces requêtes impliquent un temps de latence moyen de deux à trois secondes. L'ensemble des visualisations sont connectées à cette liste stockée localement sur le navigateur de l'utilisateur et sont actualisées en temps réel à chaque ajustement. En même temps que les articles sont récupérés à chaque chargement de la page d'accueil, les mots-clés associés aux articles sélectionnés sont également chargés localement pour permettre l'affichage du nuage de mots-clés. La page de profil fonctionne selon la même logique si ce n'est qu'il s'agit de l'historique de l'utilisateur qui est chargé au lieu des articles recommandés.

À l'heure actuelle, l'ensemble de l'interface utilisateur est uniquement adapté à une navigation sur ordinateur. Étant donné notre objectif de conception qui vise à complexifier l'interface d'un système de recommandation, il est particulièrement difficile d'implémenter ces outils dans des

terminaux dont la structure même implique une simplification de l'interface pour qu'elle puisse tenir dans les dimensions de l'écran. Adapter notre projet à ce genre d'écrans implique donc un travail conséquent qu'il ne fut pas possible d'inclure dans le planning de cette recherche. Comme solution de dépannage, nous avons désactivé l'ajustement automatique des composants de l'interface sur petits écrans (formats smartphone et tablettes) de sorte à imposer une zone d'affichage identique à ce qu'il est possible d'avoir sur un écran d'ordinateur tant que la résolution de l'écran correspond. L'utilisateur doit donc zoomer sur les composants pour pouvoir les discerner clairement mais il n'y a pas de chevauchement visuel des éléments graphiques ce qui empêcherait toute forme de navigation.

PROTOCOLE EXPÉRIMENTAL

Étant donné la nature exploratoire de cette expérimentation, nous avons surtout cherché à garantir l'engagement des participants avec la plateforme expérimentale afin de nous assurer un seuil d'interaction minimal suffisant pour disposer de données exploitables. Les résultats que nous présenterons dans le chapitre 4 ne sont donc pas généralisables mais nous donnent malgré tout une première idée de l'effet des fonctionnalités proposées. Nous avons donc sélectionné 26 étudiants, de 18 à 26 ans, issus de formations en sciences de l'information et de la communication. Étant donné leur formation ainsi que leur âge, nous cherchions à constituer une cohorte d'individus habitués à l'usage d'algorithmes de recommandation dans leur vie quotidienne et dotés d'un niveau de base d'éducation aux médias. De plus, étant donné que certains cours de leur formation demandent un suivi régulier de l'actualité, cela nous a permis d'assurer un intérêt initial pour l'actualité et dès lors, une plus grande propension à naviguer sur le site expérimental. En effet, étant donné le bagage de la plupart des participants, ceux-ci sont déjà sensibilisés à la qualité de leurs pratiques informationnelles ainsi qu'à la problématique des algorithmes de recommandation.

La plupart des méthodes expérimentales qui visent à évaluer un artefact technique sont principalement consacrées à l'évaluation de l'utilisabilité du système. Dans ce contexte, la manipulation de l'objet évalué est réalisée dans un laboratoire et dans le cadre d'une activité clairement délimitée. Il est généralement demandé aux participants de réaliser une série de tâches précises pour s'assurer que les outils étudiés soient bien mobilisés pendant l'entretien. Dans notre cas, nous nous sommes surtout intéressés à la façon dont le participant va s'approprier l'outil et l'intégrer dans ses pratiques. Cela signifie que chaque participant doit disposer d'un temps

d'adaptation plus ou moins important pour se familiariser avec le prototype. Dès lors, nous avons décidé d'adapter cette méthode en donnant accès à Alveho à distance, depuis un terminal personnel du participant et sur une période de 5 semaines. De plus, que ce soit pendant la période d'utilisation du site ou pendant l'entretien final, il n'était pas demandé aux participants de réaliser une tâche particulière. La durée de l'étude fut déterminée par des contraintes logistiques plus que par un choix méthodologique particulier. Toutefois, comme nous l'avons déjà signalé dans la section précédente, nous n'avons pas eu l'opportunité d'adapter l'interface du site pour le rendre compatible avec des écrans de petite taille (type smartphone ou tablette). Il a donc été expressément demandé aux participants de l'étude de ne pas naviguer sur ce site depuis ce genre d'appareils²⁶. Nous reviendrons sur ce point dans le cadre de notre retour critique sur la méthodologie.

Pour garantir un taux de rétention élevé sur toute la durée de l'expérimentation, les participants reçurent une compensation financière à condition de respecter plusieurs conditions de participation. Pour valider leur participation, chaque participant devait passer 20 minutes par jour sur le site, 5 jours par semaine, et ce pour chaque semaine de l'expérimentation. L'évaluation du suivi était réalisée une fois par semaine et un mail était envoyé à chaque participant qui n'atteignait pas son quota. Bien que cette information ne leur était pas communiquée au début de l'étude, chaque participant a eu droit à une seconde chance s'ils n'atteignaient pas le quota demandé pour une semaine. Malgré tout, 3 participants n'ont pas maintenu une activité suffisante pour participer à l'entretien final. Les résultats présentés dans le chapitre suivant reposent donc sur une cohorte de 23 participants.

Les données quantitatives sont issues de l'enregistrement de logs d'interaction avec la plateforme. Les participants de l'étude furent prévenus que ces informations seraient enregistrées pendant la durée de l'étude. Comme nous l'avons détaillé au tout début de ce chapitre, ces données à elles seules ne peuvent suffire pour dessiner une image précise des comportements des participants. Nous avons donc tenté d'enrichir les données récoltées par deux approches différentes : une séance d'oculométrie et un entretien semi-directif par participant.

²⁶ Cette consigne n'a pas été pleinement respectée.

À tour de rôle, pendant les 5 semaines de l'étude, chaque participant fut invité à réaliser une séance de navigation sur Alveho depuis un ordinateur équipé d'un oculomètre GazePoint GP3 HD 150Hz. La captation du mouvement des yeux de l'utilisateur vise à évaluer la place qu'occupent les visualisations de données dans le parcours de l'utilisateur. En effet, étant donné que ces composants ne demandent pas aux utilisateurs de cliquer, les logs du site ne nous donnent aucune indication quant à la consultation de ces éléments. Ces rencontres nous permirent également de maintenir le contact avec chaque participant. Nous avons envisagé de poursuivre la session par un entretien et de leur faire commenter un extrait vidéo de leur séance de navigation réalisée avec l'oculomètre. Toutefois, recourir à cette méthode portait le risque d'une surrationalisation de leurs pratiques (Eger *et al.*, 2007). Nous avons également envisagé de l'intégrer lors du second entretien mais l'obligation de réaliser une tâche de navigation ainsi que de reVISIONNER la séance auraient conduit à un allongement démesuré de la rencontre et à la baisse de qualité de l'échange. Nous avons dès lors préféré valoriser les données issues de ces sessions dans la préparation des entretiens individuels qui suivirent la réalisation de l'étude.

Nous avons également réalisé un entretien semi-directif avec chaque participant après les 5 semaines d'expérimentation. Étant donné la petite taille de notre groupe de participants, nous avons pu réaliser une première analyse manuelle des données pour dresser un profil général de chaque participant. Cette « carte d'identité » reprend le nombre de clics total par fonctionnalité du site ainsi que plusieurs graphiques qui retracent l'évolution du nombre de clics par fonctionnalité au fil du temps. Ce profil, réalisé la veille de chaque entretien, nous permet d'identifier certains phénomènes pour lesquels il apparaît nécessaire d'obtenir plus d'explications et nous permet ainsi de guider l'entretien de façon plus efficace.

L'entretien fut divisé en deux parties. Pour la première, nous avons posé plusieurs questions nous permettant d'évaluer le niveau d'aisance en informatique du participant ainsi que ses connaissances préalables à l'égard des systèmes de recommandation. Nous avons également cherché à évaluer leur rapport à l'actualité et la fréquence avec laquelle ils consultent des contenus de presse. Pour chaque thématique, le tableau suivant détaille quelques questions types :

Thématique	Questions de relance
Pratiques médiatiques (liées à l'actualité)	Est-ce que tu suis l'actualité ? À quelle fréquence ? Quel type d'actualité est-ce que tu suis ? Quelles sont tes sources d'information principales ? Sur quel support est-ce que tu suis l'actualité ?
Niveau en informatique	Quels sont les dispositifs que tu utilises le plus ? (smartphone, ordinateur portable/fixe, tablette etc.) Qu'est-ce que tu fais avec au quotidien ? Est-ce que tu as déjà eu des formations/cours portant sur l'informatique ?
Algorithmes de recommandation	Est-ce que tu sais ce qu'est un algorithme de recommandation ? Que sais-tu à ce sujet ? Peux-tu me donner des exemples de sites/applications qui en utilisent ? Lesquels utilises-tu ?

Figure 22 — Guide de la première partie de l'entretien

Une fois que nous avons récolté assez d'informations pour dresser un profil général du participant, nous l'invitons à se tourner vers un ordinateur présent dans la salle et sur lequel son compte Alveho est ouvert. Pour la suite de l'entretien, nous avons eu recours à la « walkthrough method » telle qu'elle est proposée par Light *et al.* (2018) dans le cadre de l'analyse d'applications numériques. Elle implique de parcourir de façon méthodique chaque fonctionnalité du système pour identifier l'interaction entre les caractéristiques techniques du système et ses dimensions socio-économiques. Si cette méthode est habituellement utilisée pour systématiser l'évaluation de l'interface d'une application, la « promenade » peut toutefois être dirigée par un utilisateur pour observer comment

celui-ci se conforme, ou non, aux usages prescrits par l'application. En ce qui nous concerne, nous avons demandé à nos participants de parcourir l'interface d'Alveho et de nous présenter leur routine habituelle sur le site. Lorsqu'il nous semble que le participant passe trop vite en revue certaines tâches qu'il réalise, nous lui demandons d'expliquer plus précisément les manipulations qu'il réalise pour atteindre ses objectifs. Chaque fois que c'est possible, quand un participant s'arrête sur un composant particulier du site, il lui est demandé d'expliquer à haute voix son fonctionnement. En fonction des informations récoltées grâce à sa « carte d'identité » et des propos du participant pendant l'entretien, nous orientons nos questions pour identifier les difficultés rencontrées ou pour creuser plus en profondeur certains comportements qui pourraient indiquer une forme d'autonomie. Si nous constatons qu'un composant n'a pas été utilisé, nous demandons au participant s'il est capable de fournir une explication. À la fin de la visite guidée, nous avons également demandé aux participants de modifier leurs paramètres de sorte à atteindre un objectif imposé par le chercheur. Cet objectif était relatif à l'état de leur fil d'actualité et impliquait d'inverser la tendance générale de leurs recommandations (par exemple, passer de recommandations plutôt culturelles à des recommandations sportives). En nous inspirant de la méthode d'entretien par explication (Vermersch, 2011), nous leur demandions alors d'explicitier les détails des opérations qu'ils réalisent. Le son des entretiens fut enregistré pour permettre leurs retranscriptions. Celles-ci sont accessibles dans les annexes.

Notre méthode d'analyse repose sur un étiquetage thématique (Lejeune, 2014). Un premier étiquetage déductif fut réalisé dans un premier temps sur base de la typologie des fonctionnalités intégrées au site (contrôle de l'input, de l'output, du processus, visualisations de données et outils d'aide). Étant donné les difficultés rencontrées par la majorité de nos participants (cf. chapitre 4), nous avons affiné notre analyse par un étiquetage inductif centré sur les problèmes rencontrés par nos participants. Le tableau qui reprend les thématiques identifiées ainsi que des verbatims correspondants est disponible dans les annexes. Étant donné la non-représentativité de notre échantillon, lors de notre analyse, nous ne nous sommes pas attardés sur un décompte quantitatif. À l'exception de certaines tendances globales pour lesquelles nous précisons si le comportement observé correspond à l'ensemble de l'échantillon, la majorité de l'analyse vise principalement à identifier un panel de comportements ou d'affects et à retracer leurs causes. L'ensemble des résultats issus de cette étude sont détaillés dans le prochain chapitre.

CHAPITRE 4 — RÉSULTATS

Après avoir présenté notre protocole méthodologique, nous arrivons progressivement au terme de cette recherche. À travers le développement d'Alveho et son évaluation en conditions réelles, nous avons cherché à identifier des modalités d'interaction humain-machine qui puissent encourager l'utilisateur à développer une forme de réflexivité critique sur les effets de ces dispositifs sur ses usages. Grâce à notre analyse critique des préoccupations sociales relatives à l'intégration de ces technologies dans notre écosystème informationnel, nous en sommes arrivé à situer une problématique de design importante autour de la question de l'autonomie. À travers l'exploration des travaux de Foucault sur le sujet et son rapport au pouvoir et à la norme, nous en sommes arrivé à formaliser l'autonomie comme une compétence créative distribuée qui repose sur la capacité de l'individu à moduler les relations de pouvoir dans lesquelles il s'investit. Nous avons ainsi postulé que fournir à l'utilisateur des outils de contrôle sur l'algorithme peut encourager le développement de cette autonomie en facilitant l'exploration des recommandations et en permettant à l'utilisateur de relativiser la pertinence des choix proposés par l'algorithme.

Ce quatrième chapitre est consacré à la présentation de nos résultats et à leur discussion. Nous développerons notre analyse des résultats en trois niveaux successifs. Tout d'abord, nous présenterons une vue d'ensemble des comportements observés sur Alveho ainsi qu'une première analyse des commentaires issus de nos entretiens. Ce premier niveau d'interprétation nous permettra de dégager une vue d'ensemble des comportements qui furent adoptés sur la plateforme par nos participants. Nous reviendrons notamment sur les données issues des logs enregistrés dans la base de données.

Dans un deuxième temps, nous reviendrons en détail sur chacune de nos hypothèses. Étant donné les nombreuses difficultés rencontrées par nos participants, nos résultats apparaissent beaucoup plus nuancés que nous ne l'avions escompté. Nous profiterons de ce constat pour identifier de façon précise les éléments qui ont conduit certains utilisateurs à ne pas utiliser les outils mis à leur disposition. Nous saisisons également cette occasion pour localiser les modalités de contrôle de l'algorithme qui furent des leviers de développement de l'autonomie de nos participants.

Enfin, sur base des pistes identifiées grâce à ces deux premiers niveaux d'analyse, nous dessinerons plusieurs perspectives de recherche. Nous insisterons notamment sur deux leviers d'autolégitimation qui nous apparaissent comme prometteurs : l'anthropomorphisation de l'algorithme ainsi que l'interruption de l'utilisateur. Pour conclure ce chapitre, en complément de notre analyse principale, nous présenterons également une activité d'éducation aux médias qui fut réalisée avec l'aide d'Alveho. Ce sera pour nous l'occasion de mettre en perspective nos résultats précédents à travers l'usage d'Alveho dans un contexte à la fois contrôlé, mais également collectif.

RÉSULTATS

Évaluation générale de l'activité

Nos entretiens nous permettent de confirmer que 21 de nos 23 participants disposaient de connaissances préalables relatives à l'usage de systèmes de recommandation. Même s'ils n'étaient pas capables d'expliquer des détails techniques complexes, ils pouvaient nommer plusieurs services commerciaux mobilisant ce genre d'outils et décrire leurs principes de fonctionnement de base. Ils étaient également conscients des enjeux éthiques qui gravitent autour de ces systèmes. Plusieurs d'entre eux ont d'ailleurs spontanément mentionné la notion de bulle de filtre. Quatre participants avaient une expérience préalable en programmation informatique. Parmi ces participants, l'un d'entre eux a suivi des études supérieures en informatique. Pour tous les autres, leur expérience quotidienne avec un ordinateur se limitait à l'usage d'outils de bureautique dans le cadre d'activités récréatives ou de leurs cours à l'université.

Pour rappel, nous avons séparé les fonctionnalités de contrôle de l'algorithme en trois catégories en fonction de l'étape de la recommandation sur laquelle elles agissent : *input* pour le contrôle des données fournies à l'algorithme, *output* pour le contrôle des recommandations affichées, *processus* pour le contrôle de la logique de l'algorithme. Le contrôle des inputs est accessible depuis la page de profil et permet de masquer certains éléments de l'historique à l'algorithme. Le contrôle du processus se fait en ajustant les curseurs disponibles sur la page principale du site. Enfin, le contrôle des outputs se fait également sur la page principale en sélectionnant les catégories que l'on souhaite afficher.

Selon les données de navigation récoltées, la plupart des fonctionnalités proposées à l'utilisateur furent utilisées (voir figure 23). Toutefois, les interviews réalisées à la suite de l'expérimentation nous indiquent que la plupart des interactions enregistrées pour les contrôles de type input furent

produites par une poignée d'utilisateurs qui cliquèrent sur ces fonctionnalités de façon aléatoire sans nécessairement comprendre leur fonctionnement. Si nous prenons en compte ce constat, il en ressort qu'aucun de nos participants n'a utilisé cette fonctionnalité volontairement. Ce résultat contraste fortement avec les résultats d'Harambam *et al.* (2019) dont les participants plébiscitent plus le contrôle de l'input que de l'output. Seuls 13 participants utilisèrent les outils de contrôle de l'algorithme, ce qui signifie que 10 participants n'ont jamais changé leurs paramètres à l'exception des outils de contrôle de type output qui furent utilisés par l'ensemble des participants. La plupart des interactions se déroulent lors des quelques premiers jours même si le taux d'utilisation des contrôles de l'output reste élevé tout au long de l'étude. Ce constat correspond aux observations d'Ekstrand *et al.* (2015). Seulement 7 utilisateurs ont interagi avec les outils de contrôle de l'algorithme au-delà de la première semaine de l'étude. Dans l'ensemble, le niveau d'activité avec le système reste faible.

L'analyse des vidéos issues de la séance de navigation avec oculomètre confirme ce constat. Quel que soit le moment de l'étude pendant lequel la séance d'oculométrie s'est déroulée, la quasi-totalité des participants se contenta de dérouler le fil d'actualité. Le suivi des yeux indiqua que le nuage de mots ainsi que le graphe circulaire n'étaient que très peu consultés. Au-delà de ces quelques constats, l'usage de l'oculomètre ne fut pas conclusif. Nous n'avons pas pu observer l'usage des outils de contrôle lors de ces séances ce qui implique que les données récoltées ne nous apportent pas d'informations complémentaires pertinentes.

Étant donné les faibles résultats issus de cette analyse qualitative de l'activité générale des participants, nous n'avons pas cherché à quantifier nos observations. Une exception notable fut la séance organisée avec un étudiant ayant un bagage en sciences informatiques. Celui-ci utilisa l'outil de recherche pour contourner les recommandations et se renseigner sur des thématiques précises.

Les participants qui utilisèrent les outils de contrôle de l'algorithme étaient plus susceptibles d'avoir un plus haut niveau de littératie informatique ou de développer des usages plus complexes lorsqu'ils naviguent sur des systèmes de recommandation commerciaux. Toutefois, un échantillon plus représentatif serait nécessaire pour confirmer ces observations. Il est aussi important de noter que les utilisateurs qui interagirent avec le système n'étaient pas toujours certains des effets de leurs actions.

Enfin, les outils d'aide (pour rappel : une infobulle pour chaque composant, une visite guidée à la première connexion et un manuel) furent ignorés par l'ensemble des participants. Si les logs indiquent que plusieurs participants se sont rendus sur le manuel du site, ceux-ci nous ont signalé qu'ils n'en ont pas vraiment lu le contenu. À chaque fois qu'ils furent interrogés sur leur utilisation de ces outils d'aide, ils ont tous déclaré qu'ils n'en sentaient pas le besoin ou que cela ne les intéressait pas. Plusieurs participants déclarèrent que le site leur semblait intuitif.

Dans l'ensemble, nos observations indiquent que les outils de contrôle furent plus difficiles à comprendre que nous ne l'avions estimé au départ. 10 utilisateurs n'ont jamais touché aux contrôles et sont restés sur leurs paramètres par défaut lors de l'expérimentation. Parmi les 13 participants qui utilisèrent les mécanismes de contrôle, seulement 3 sont parvenus à expliquer le raisonnement qui les a amenés à agir sur ces composants. Il s'agissait principalement de modifier la liste des recommandations pour vérifier qu'ils ne rataient pas de contenus qui pourraient les intéresser. Aucun participant n'est toutefois parvenu à expliquer les deux paramètres et leur rapport avec l'algorithme.

En ce qui concerne l'utilisation du filtrage des outputs, l'objectif des participants fut de réduire leur fil d'actualité à des intérêts spécifiques. Certains utilisateurs l'ont tellement filtré qu'il n'y avait plus assez d'articles pour justifier l'usage d'un algorithme de recommandation. Notre analyse de l'évolution des logs de chaque participant nous confirme qu'ils s'arrêtent très rapidement sur une configuration spécifique qu'ils conservent pour le reste de l'expérimentation, à l'exception des outils de contrôle des outputs : « Quand j'ai trouvé cet équilibre-là, j'ai plus vraiment trifouillé dans tout ça ». (Entretien 3, p.5) C'est souvent une satisfaction avec l'algorithme qui conduit à cette stabilisation :

Bah le reste de l'algorithme me plaisait. Et je voulais pas trop déséquilibrer l'algorithme actuel, enfin je voulais juste plus de culture. (Entretien 13, p.8)

Le contrôle de l'algorithme je n'ai pas, j'ai laissé ça comme ça parce que ça me convenait. (Entretien 2, p.3)

Enfin, lors de la mise en situation où il leur fut demandé de modifier leur fil d'actualité, aucun participant ne parvint à réaliser la tâche demandée sans recourir au filtrage des outputs.

Logs distribution over time

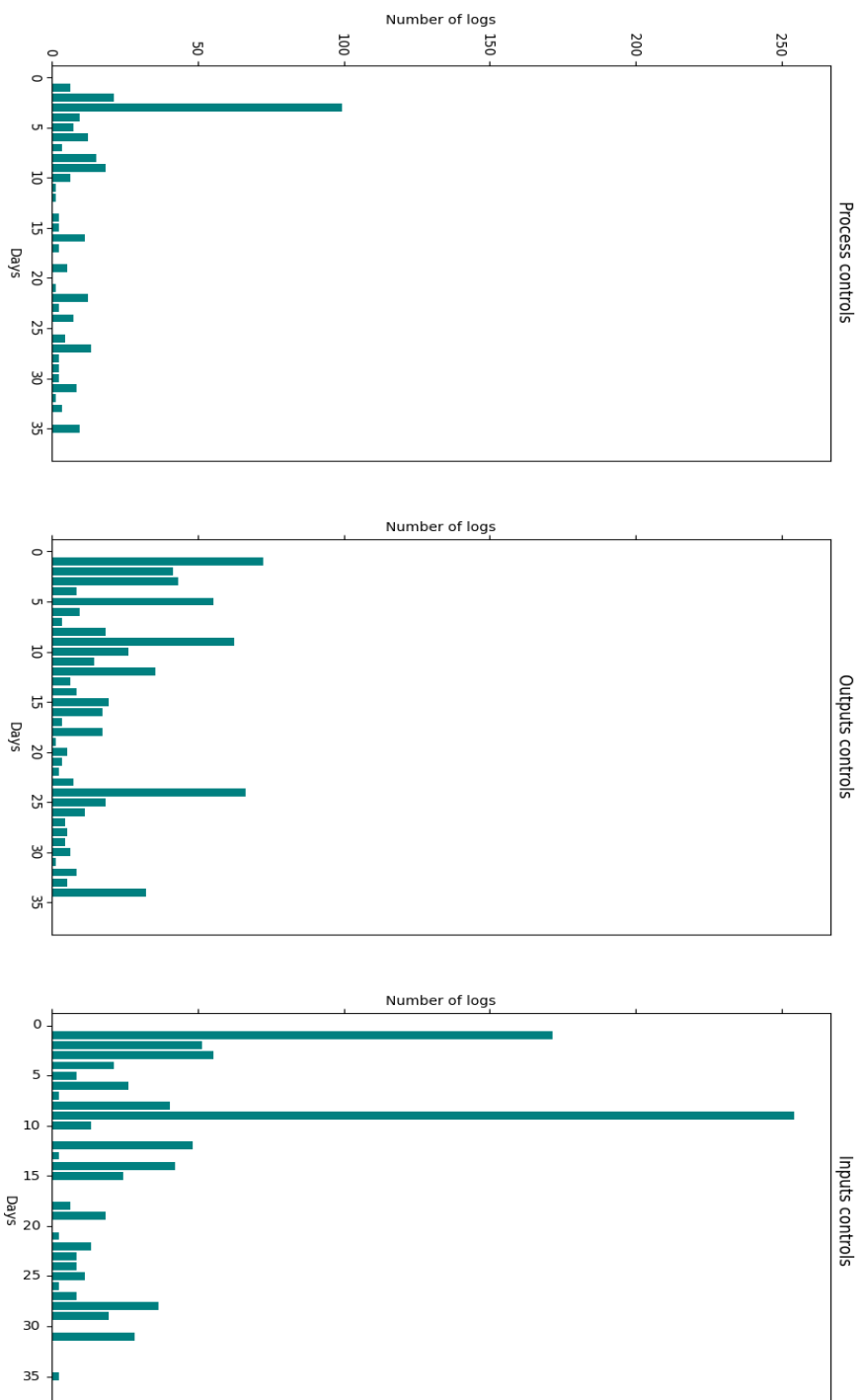


Figure 23 — Évolution du nombre de logs enregistrés pour chaque type de contrôle. « Process controls » réfère aux outils de contrôle de l' algorithme, « Outputs controls » au filtrage des résultats tandis que « Inputs controls » réfère à la gestion de l' historique.

Compréhension des répercussions

La tâche qui fut la plus compliquée à réaliser par nos participants fut l'identification de l'effet de leurs actions sur leurs recommandations, surtout lorsqu'ils interagirent avec les contrôles associés aux inputs et à l'algorithme. Toutefois, ils purent parfaitement identifier l'effet du filtrage au niveau des outputs. Ce constat peut être expliqué par la proximité des composants de contrôle au fil d'actualité et à l'actualisation immédiate de ce fil. Tandis que le filtrage de l'output se déroulait sur la page principale, le filtrage par input demandait de se rendre sur une page secondaire sur laquelle le fil d'actualité n'était pas directement accessible. Dans le cas du contrôle direct de l'algorithme, bien que le fil d'actualité soit adjacent au composant, il est parfois difficile de constater une différence immédiate quand les premiers résultats restent identiques. Bien que cela puisse varier en fonction de la taille de l'écran, seuls les 4 premiers articles sont visibles sans devoir faire défiler le fil d'actualité. Si les changements au sein d'un ensemble de recommandations se situent au-delà de ces 4 premiers articles, il est donc nécessaire de faire défiler l'écran pour s'en rendre compte.

En ce qui concerne les poids de l'algorithme, plusieurs participants ont arrêté de les ajuster parce qu'ils ne pouvaient pas voir de différences dans leur fil d'actualité. Si leur première manipulation ne conduit pas à des changements notables, le composant est ignoré pour la suite de l'étude : « J'ai fait ça et j'ai vu que ça ne changeait pas grand-chose. Enfin, en tout cas, c'est l'impression que j'ai eu. Je l'ai pas utilisé par après. » (Entretien 23, p.4) Selon un autre participant : « J'avais essayé, mais pareil, je ne voyais pas trop ce que ça changeait. Du coup, je n'ai pas touché. » (Entretien 20, p.4). Ces propos sont corroborés par les traces de navigation qui montrent clairement que la majorité des interactions sont réalisées lors des dix premiers jours.

Plutôt que d'aider les utilisateurs à observer les effets de leurs actions, la présence de visualisations de données s'est avérée une source de confusion pour certains de nos participants : « Il y avait beaucoup de couleur, ça me paraissait très flou. » (Entretien 15, p.4). Plusieurs utilisateurs ne furent pas capables de comprendre le graphique en zones empilées. De plus, étant donné que nous avons utilisé le même type de graphique pour représenter la distribution du fil d'actualité et de l'historique de navigation, plusieurs participants ont cru que ces deux graphiques représentaient la même chose.

Il apparaît également que l'interactivité de ces composants a contribué à cette confusion générale. Un de nos participants nous signale ainsi que : « J'ai arrêté d'y toucher parce que quand ça change

beaucoup on a l'impression d'avoir un contenu totalement différent et du coup ça donne moins envie au départ. Parce qu'on a envie d'avoir un truc stable et de savoir à quoi on s'attend » (Entretien 13, p.9).

L'interface comprenait également un panel coulissant permettant de comparer la distribution de son fil d'actualité avec la distribution d'un fil non recommandé. Une fois activé, ce panel vient se placer au-dessus du fil d'actualité. Étant donné qu'il vient s'ajouter aux visualisations, l'un de nos participants nous a confié son malaise à l'égard de cette visualisation : « Je sais plus comment ça marchait, mais je sais que je me suis dit c'est quoi ça ? Pas incroyable quoi » (Entretien 2, p.4).

Perception des composants

Dans l'ensemble, le volume d'interactions avec les fonctionnalités de contrôle d'Alveho est resté particulièrement faible tout au long de l'étude. Lorsque nous avons interrogé nos participants à ce sujet, plusieurs d'entre eux nous ont signalé qu'ils n'avaient pas vu qu'il était possible d'interagir avec ces éléments. Les utilisateurs ayant vu ces outils mais ne les ayant pas utilisés nous ont signalé que l'absence d'interaction était liée au fait qu'ils n'étaient pas habitués à ce genre de mécanisme. L'un de nos participants a reconnu qu'« On nous propose pas ce service-là dans les réseaux sociaux ou même les médias numériques en général. C'est pour ça que je me suis pas posé la question, ça me paraissait pas évident. » (Entretien 4, p.9) ». Un autre participant nous signale : « Je n'ai même pas essayé et je ne saurais même pas dire pourquoi. J'imagine que ça m'a pas... Bah ça me parlait pas » (Entretien 5, p.3).

Toutefois, plusieurs participants ont tenté d'influencer l'algorithme indirectement en cliquant sur des articles spécifiques pour amplifier certains aspects de leur profil. Ce comportement a déjà été observé sur d'autres réseaux sociaux numériques comme tentative de contournement du système de recommandation qui ne propose pas de modalités d'interactions plus directes (Wiard *et al.*, 2022). Dans le cas de cette étude, il est dès lors apparu plus intuitif pour nos participants d'appliquer des tactiques déjà éprouvées ailleurs plutôt que d'apprendre à utiliser les fonctionnalités mises à disposition, même si celles-ci permettent d'atteindre le même objectif de façon plus précise.

Interruption du flux de l'activité

Lors de l'étude, nous avons pu constater que la majorité des comportements de navigation gravitent autour d'un défilement, plus ou moins long, du fil d'actualité. Malgré le faible volume

d'interactions avec le système, il nous est possible d'identifier plusieurs occasions pendant lesquelles nos participants furent amenés à prendre du recul sur leurs actions et à interrompre le flux de leur lecture.

Deux participants de l'étude, se connaissant en dehors du cadre de l'étude, se sont retrouvés à naviguer sur Alveho au même moment tout en se trouvant l'un à côté de l'autre. Nous n'avions pas implémenté une fonctionnalité permettant aux utilisateurs de comparer leur fil d'actualité avec celui de quelqu'un d'autre. Toutefois, cette situation leur a permis d'observer les différences entre leurs fils d'actualité en regardant sur leurs écrans respectifs. Pour l'un d'entre deux, après avoir comparé son fil d'actualité avec celui de sa connaissance, il constate que leurs deux fils d'actualité divergent totalement. Ce constat l'a ensuite encouragé à manipuler les paramètres de l'algorithme pour obtenir un résultat différent.

Si nous n'avions pas implémenté un outil de comparaison, nous avions par contre implémenté un panneau coulissant qui permettait de comparer les distributions entre le fil d'actualité de l'utilisateur et un fil d'actualité non recommandé. Toutefois, cet outil n'a pas été utilisé. Probablement car il était désactivé par défaut et qu'une fois activé, il obstruait la vue sur le fil d'articles. Un seul de nos participants, avec un bagage préalable en informatique, a eu recours de façon régulière à cet outil pour évaluer la nature globale de son fil d'actualité au début de sa séance de navigation.

La taille de la liste de recommandation semble également avoir influencé le comportement de certains de nos participants. Nous avons fixé une limite de 100 articles pour des raisons techniques. Notre algorithme prenait en compte plusieurs milliers d'articles issus de notre base de données pour faire produire une liste, mais seuls les 100 premiers apparaissaient sur la page du site. Modifier le filtrage de l'output cachait simplement certains éléments de cette liste tandis que changer les paramètres de l'input ou de l'algorithme enclenchait un rafraîchissement complet de cette liste. Nous avons initialement prévu que nos participants réaliseraient plusieurs microsessions de navigation par jour et que, par conséquent, ils n'atteindraient pas la fin de la liste. Toutefois, il s'est révélé que la majorité d'entre eux réalisaient leur navigation quotidienne en une seule grande

séance²⁷. Dès lors, il était fréquent pour certains de nos participants d'arriver au bout de leur liste. Lorsque cela arrivait, ils revenaient alors sur les outils de contrôle pour les modifier et générer une nouvelle liste afin de vérifier s'ils n'avaient pas raté certains contenus.

Enfin, la présence de statistiques personnalisées semble affecter la façon dont nos participants interprètent leur comportement et se situent au sein de la communauté de participants. La simple présence de statistiques très simples comme la moyenne du nombre d'articles lu par les autres participants a conduit plusieurs d'entre eux à s'interroger sur la signification de ce chiffre et à y réagir de façon plus ou moins consciente !

C'était informatif, mais dans ma tête c'était aussi un petit challenge, je me disais « il faut que je lise, que je sois dans la moyenne. (Entretien 20, p.4)

C'est vrai que ça me perturbait peut-être un peu de voir le nombre d'articles que d'autres gens lisaient en moyenne [...] Souvent je pense, chez moi, il était un peu en dessous de la moyenne. [...] Mais je ne sais pas, c'est comme si je me mettais cette pression où je me disais « zut, je ne lis pas assez ». (Entretien 12, p.6)

DISCUSSION DES RÉSULTATS

Retour aux hypothèses

Avant de préciser l'analyse de nos résultats, nous proposons de revenir dans un premier temps à nos six hypothèses.

H1 : La présence de fonctionnalités de contrôle de l'algorithme encourage l'utilisateur à interagir avec le système. Les outils de filtrage des outputs furent spontanément utilisés par les participants de l'étude. Toutefois, ce ne fut pas le cas du contrôle de l'input et de l'algorithme. Si dans le cas de l'input, ces difficultés peuvent être associées à des problèmes d'utilisabilité, il est plus difficile d'expliquer pourquoi le contrôle de l'algorithme fut moins mobilisé.

H2: Contrôler le fonctionnement de l'algorithme de recommandation facilite la compréhension du fonctionnement du système. Étant donné le faible volume d'interactions avec le système, nous ne disposons pas d'assez de données nous permettant de valider ou d'invalidier cette hypothèse.

²⁷ Nous reviendrons sur ce point ci-dessous dans la section consacrée aux limitations de l'étude.

Toutefois, étant donné que la totalité de nos participants n'était pas capable d'expliquer comment l'algorithme fonctionnait avec précision, les résultats en notre possession semblent indiquer que le contrôle n'est pas suffisant pour garantir une meilleure compréhension des mécanismes de recommandation.

H3 : Comprendre le fonctionnement du système encourage l'utilisateur à développer un jugement critique sur les effets de la recommandation sur ses pratiques. Bien que nos participants ne comprennent pas précisément comment fonctionnait l'algorithme implémenté sur Alveho, il est important de souligner que la majorité de nos participants connaissaient les enjeux médiatiques généraux relatifs à la recommandation automatique d'informations étant donné leur parcours universitaire. Toutefois, ce n'est pas pour autant qu'ils ont développé un jugement critique relatif à l'emploi d'Alveho. Il nous semble donc tout à fait possible d'invalider cette hypothèse d'une relation entre une connaissance critique et le développement d'un jugement critique relatif à ses propres pratiques.

H4 : Contrôler le fonctionnement de l'algorithme de recommandation encourage l'exploration de recommandations alternatives. L'usage conscient d'outils de contrôle semble associé à une stratégie d'exploration. Toutefois, plusieurs conditions complémentaires semblent nécessaires pour initier cette exploration. Nous explorerons plusieurs pistes explicatives dans la suite de ce chapitre.

H5 : L'exploration de recommandations alternatives encourage l'utilisateur à développer un jugement critique sur les effets de la recommandation sur ses pratiques. Nos données ne nous permettent pas de valider ou d'invalider cette hypothèse. Toutefois, nous avons pu identifier certains leviers d'action qui semblent avoir un effet positif sur l'exercice d'une forme de réflexivité.

H6 : L'exploration de recommandations alternatives encourage l'utilisateur à développer une stratégie d'intégration de la recommandation automatisée dans ses pratiques. Nos données ne nous permettent pas de valider ou d'invalider cette hypothèse. Nos participants ont cumulé l'usage d'Alveho à leurs pratiques médiatiques habituelles de telle sorte que nous n'avons pas pu identifier une stratégie particulière d'intégration de ce système à leurs écosystèmes informationnels.

Selon les résultats de l'étude réalisée par Harambam *et al.* (2019), l'intégration d'outils de contrôle de l'algorithme ainsi qu'un profil utilisateur intelligible sont deux fonctionnalités qui seraient valorisées et désirées par les utilisateurs d'algorithmes de recommandation d'actualités. Lors de

l'entretien final, plusieurs de nos participants nous ont également signalé apprécier le principe général d'une contrôlabilité d'un algorithme de recommandation. Il semble toutefois qu'il existe un fossé non négligeable entre ces désirs et les comportements réellement mis en œuvre. Bien que nos résultats ne nous permettent pas de répondre avec certitude à l'ensemble de nos hypothèses, ils nous permettent toutefois de constater que la présence d'outils de contrôle et leur manipulation ne suffit pas, seule, à faciliter la compréhension des mécanismes ou à favoriser une prise de recul sur ces pratiques nécessaires au développement d'un usage autonome. La suite de notre analyse est consacrée à l'identification des conditions complémentaires nécessaires au développement de cet usage mais aussi des facteurs qui peuvent interférer avec cette dynamique.

Freins potentiels à l'autonomie

Transparence

Nos résultats mettent en évidence toute l'ambiguïté et la difficulté d'un design qui se voudrait « transparent ». En ayant recours à un algorithme prenant la forme d'une combinaison linéaire et en traduisant cette combinaison par une série de curseurs permettant d'ajuster les variables de l'équation, nous avons ainsi l'impression, très naïvement, de proposer un design qui soit plus « honnête » qui donne un accès plus « direct » à la réalité technique sous-jacente. Or, il s'est avéré que ce sentiment de proximité entre l'interface et la mécanique algorithmique est relatif au profil de l'utilisateur. Ceci corrobore l'argument que nous avons développé dans le chapitre 2 selon lequel il ne peut exister une interface qui serait, par nature, transparente. Harambam *et al.* (2018) adoptent une démarche inverse et suggèrent de recourir à des personae pour introduire l'utilisateur à la configuration du système. Nous explorons cette piste dans la section consacrée aux perspectives de recherche ouvertes par notre étude.

On pourrait toutefois nous objecter que cette difficulté soit plutôt liée à des problèmes d'utilisabilité de notre système. S'il était effectivement nécessaire de confirmer les paramètres pour initier l'actualisation du flux et qu'un délai était parfois nécessaire avant l'apparition de nouvelles recommandations, les composants étaient bien labellisés, proches du fil d'actualité et accompagnés de plusieurs modules informatifs. Ce n'était pas le cas pour le filtrage des inputs et pour lequel nous reconnaissons plus volontiers un problème d'utilisabilité plutôt qu'une difficulté intrinsèque à la méthode d'interaction.

Une difficulté similaire concerne les visualisations de données mobilisées sur les deux pages principales du site. Si le graphique circulaire de la page principale fut facilement interprété, le graphique en zones empilées a posé plus de problèmes. Incorporer des graphiques peu communs implique donc de soigneusement prendre en compte le niveau de littératie visuelle nécessaire (Boy *et al.*, 2014).

Gestion de la charge cognitive

Il est évident que le recours à une interface complexe conduit à une charge cognitive plus importante pour l'utilisateur. Dans notre cas, il nous semble possible de fournir un diagnostic plus précis de cette charge cognitive en recourant à la notion de « map shock » (Blankenship et Dansereau, 2000). À l'origine, ce concept est employé pour décrire un sentiment d'écrasement ou d'accablement face à la complexité d'une carte conceptuelle. Cette réaction génère une émotion négative qui décourage le lecteur à s'engager dans le déchiffrement des interconnexions qui composent la carte. Bien que nos participants n'eussent pas affaire à une carte conceptuelle, ils devaient traiter plusieurs visualisations de données actualisées en temps réel à chaque fois qu'ils manipulaient un bouton ou un curseur. En plus de prendre en compte les connaissances préalables nécessaires à la compréhension de certaines visualisations, il apparaît également crucial de prendre en compte une introduction progressive de l'utilisateur aux composants interactifs afin de ne pas le décourager de toute interaction future.

La force de l'habitude

Nous avons introduit, dans le premier chapitre de cette thèse, les travaux de Davis (2020) qui suggère que la façon dont un artefact influe sur nos comportements peut s'exercer avec plus ou moins d'intensité selon qu'il impose, demande, encourage, décourage, refuse ou autorise certaines actions spécifiques. Au-delà des caractéristiques techniques de l'artefact, ce sont également les conditions de perception qui déterminent la force de cette relation. Une affordance n'est pas une propriété technique mais plutôt une relation de pouvoir qui s'établit entre un artefact, un utilisateur et une situation d'usage. La demande d'un objet peut n'être qu'une simple requête pour quelqu'un d'autre. Si des paramètres par défaut d'un système peuvent n'être qu'une simple proposition pour un utilisateur aguerri, ils peuvent prendre la forme d'une imposition pour un utilisateur novice. Dans le cas de notre étude, plusieurs participants craignaient qu'en modifiant les paramètres du système, ils ne puissent pas revenir en arrière en cas de problèmes.

La perception des prises sur lesquelles l'utilisateur peut agir est dès lors dépendante de leurs expériences préalables avec d'autres technologies similaires qui agissent comme des outils de disciplinarisation (Bonenfant et Farmer, 2012). Les gestes, les comportements mais aussi les tactiques de résistance adoptés sur ces plateformes sont incorporés et transposés par leurs utilisateurs à d'autres domaines de leur écosystème médiatique.

La gestion de la charge cognitive, discutée ci-dessus, ne se limite donc pas à des considérations de stimulation sensorielle (limiter le nombre d'éléments réactifs, limiter la complexité des représentations graphiques, équilibrer le texte, etc.). Cette gestion implique également de prendre en compte les gestes que les utilisateurs vont tenter d'appliquer spontanément et les gestes qu'ils devront apprendre. En l'occurrence, le système que nous avons implémenté ne comprend pas un mécanisme permettant à l'utilisateur de spécifier son intérêt pour un contenu spécifique à l'image de « likes » de Facebook qui est devenu une fonctionnalité largement répandue. Au mieux, l'utilisateur peut cliquer sur un article pour simuler une lecture qui sera ajoutée au profil de l'utilisateur. Le recours à ces tactiques est moins efficace que les outils d'Alveho à disposition et demande un effort cognitif plus important. Toutefois, le « coût d'entrée » est plus faible.

Nous ne cherchons pas à préconiser une pratique de design qui s'aligne sur les mécaniques de quelques plateformes populaires. Il est toutefois nécessaire d'envisager le trajet par lequel il sera possible d'amener l'utilisateur à explorer des fonctionnalités qui ne correspondent pas à ses habitudes de sorte qu'il ne se limite pas aux paramètres par défaut et à quelques tactiques de résistance.

Leviers réflexifs

Interruption

La majorité des plateformes numériques plaçant un fil d'actualité au centre de leur interface ajoutent progressivement des contenus à la fin du fil au fur et à mesure que l'utilisateur fait défiler l'écran de sorte à donner l'illusion d'un flux de contenus infinis. Ce mécanisme ne demande aucune interaction de la part de l'utilisateur pour générer ces nouveaux éléments. Or, il semble que limiter la liste de recommandations peut conduire l'utilisateur à interrompre sa navigation pour s'intéresser aux outils de contrôle et générer de nouvelles recommandations. Toutefois, limiter ainsi la liste des recommandations peut, en fonction du domaine d'application, réduire la qualité de l'expérience utilisateur.

De façon plus générale, ce constat nous amène à considérer un design plus intrusif qui ne se limite pas à compenser l'asymétrie informationnelle entre l'utilisateur et le système ou à mettre à disposition des outils de contrôle mais qui va volontairement interrompre l'activité en cours. Nous explorons cette piste dans la prochaine section.

Comparaison

Nous n'avons pas de données qui nous permettent d'affirmer avec certitude que la génération de plusieurs ensembles de recommandations et de leur comparaison conduit à plus d'autonomie de la part de l'utilisateur du système. Toutefois, l'absence de cette possibilité de comparaison nuit visiblement à l'usage des outils de contrôle comme le démontre la très faible utilisation du filtrage de l'input. Si comparer plusieurs générations simultanées de recommandations n'est pas une condition suffisante, elle est au moins nécessaire pour stimuler l'utilisation des fonctionnalités de contrôle. Cela implique donc de parvenir à intégrer l'ensemble des mécanismes nécessaires dans la zone d'affichage de l'écran pour que l'utilisateur puisse observer en temps réel les changements qui découlent de ses actions. Si cela ne pose pas de difficultés majeures pour une navigation sur ordinateur, ceci ajoute une couche de complexité supplémentaire au développement d'une interface mobile.

Nos données semblent toutefois indiquer que permettre à l'utilisateur de comparer des recommandations associées à des profils différents du sien est un levier prometteur. Bien plus qu'une comparaison entre deux sets issus du profil de l'utilisateur, ce genre de comparaisons semble indiquer de façon plus insistante que cette recommandation doit faire l'objet d'une certaine forme de préoccupation.

Statistiques

Des recherches ultérieures ont déjà mis en évidence que la présence de marqueurs d'approbation sociale (nombre de likes, de partages, de commentaires etc.) influence la façon dont des utilisateurs sélectionnent les informations qu'ils consultent (Messing et Westwood, 2014).

Bien que nous n'ayons pas incorporé de tels marqueurs sociaux, il est apparu que les quelques statistiques présentes sur la page de profil ont conduit certains de nos participants à s'interroger sur leurs pratiques et à se positionner par rapport aux chiffres qui leur étaient présentés. Toutefois, étant donné qu'il ne s'agit pas d'une fonctionnalité qui fut approfondie lors de la phase de

développement, nous ne disposons pas d'assez de données pour approfondir notre analyse. Nous aborderons cette problématique dans la prochaine section.

Conclusion

Pour conclure, bien que des études plus approfondies soient nécessaires, nos résultats nous permettent d'étoffer notre triptyque agentivité-intelligibilité-légitimité présenté en conclusion du chapitre 2.

À l'égard de l'agentivité, la simple présence d'outils de contrôle ne semble pas être une condition suffisante pour garantir une prise en main de l'algorithme par l'utilisateur. Il est dès lors raisonnable de penser que ces fonctionnalités doivent être intégrées dans une réflexion plus large sur la charge cognitive globale à laquelle il est possible de soumettre l'utilisateur de sorte à garantir l'adoption progressive de nouveaux gestes.

De façon similaire, la production d'une connaissance de soi et d'un jugement critique de ces pratiques ne peut reposer sur un simple rééquilibrage de l'asymétrie informationnelle entre l'utilisateur et le système à travers la mise à disposition de graphiques et de statistiques. L'intégration de mécanismes de rupture de la routine (ex : limitations du nombre d'objets recommandés) et de système permettant de comparer ces recommandations avec celles d'autres participants s'avère à cet égard prometteur.

Enfin, les enjeux de légitimité nous semblent les plus difficiles à aborder car ils impliquent de faire en sorte que l'individu reconnaisse non seulement ses pratiques comme un sujet de préoccupation mais aussi qu'il se sente également compétent pour s'en soucier et pour s'impliquer dans la manipulation du système. Pouvoir comparer ses pratiques avec celles d'autres utilisateurs s'avère encore une fois un levier potentiel à considérer.

LIMITATIONS

Plusieurs limitations relatives à notre méthodologie doivent être soulignées avant de nous intéresser aux pistes de recherche ouvertes par nos résultats. Pour permettre aux participants de toucher une rétribution pour leur participation, nous avons décidé d'imposer un temps de navigation minimal quotidien de 20 minutes. Un indicateur sur leur page de profil passait au vert quand le temps minimal était atteint. Sachant que dans certains cas, nous sous-estimions le temps passé sur le site,

cette contrainte temporelle fut considérée trop longue en regard des habitudes de navigation de nos participants.

De plus, comme nous l'avions souligné dans le chapitre précédent, l'interface d'Alveho n'est pas suffisamment réactive pour que le site puisse être utilisé sur un écran de tablette ou de smartphone. La lecture d'articles reste possible mais les outils périphériques du flux principal sont inutilisables. Dès lors, il fut demandé aux participants de ne pas naviguer sur Alveho via ces types de terminaux. Il est toutefois évident que l'accès à l'actualité passe aujourd'hui en grande partie par une consultation sur périphérique mobile. Il ne fut dès lors pas surprenant de constater que cette consigne ne fut pas respectée par plusieurs participants qui consultaient le site lors de trajets en transports en commun. La solution la plus rapide à implémenter fut de figer le redimensionnement des composants de sorte que la fenêtre d'affichage ne s'aligne pas sur les dimensions de petits écrans.

Alors que nous comptions sur une pratique plus fragmentée et plusieurs visites quotidiennes sur le site par utilisateur, la plupart des participants nous ont confié réaliser leurs séances de navigation d'une seule traite de sorte à s'en « débarrasser » pour le reste de la journée. Étant donné ces deux contraintes, bien que nous ayons tenté de simuler des conditions réelles de navigation, l'usage d'Alveho fut plus artificiel que nous ne l'aurions souhaité.

Une autre limitation évidente concerne la taille de notre échantillon. Étant donné que nous avons préféré opter pour une démarche qualitative et un petit échantillon, nous ne pouvons pas généraliser les quelques résultats que nous avons identifiés plus haut. De plus, nos participants disposaient tous d'un niveau de connaissances non négligeable à l'égard des enjeux médiatiques des algorithmes de recommandation. Il est fort probable que nos résultats ne sont pas applicables à un public moins averti.

Enfin, notre méthode de recueil de données implique également quelques limitations notables. Outre les difficultés rencontrées lors de l'usage de l'oculomètre qui furent déjà mentionnées, l'interprétation des logs d'interaction doit être réalisée avec précaution. En effet, notre définition des logs à enregistrer est un choix partial qui peut être contesté. Certains comportements (comme le passage sur une info-bulle par exemple) n'ont pas été enregistrés. Ces choix découlent souvent de considérations techniques et pragmatiques afin d'optimiser le temps de développement disponible. Cela signifie donc que l'image de l'utilisateur obtenue par cette méthode est également

incomplète. La réalisation d'entretiens semi-directifs nous permet de mitiger partiellement ce problème. Toutefois, la réalisation par le participant d'une visite guidée du prototype après les 5 semaines de l'étude s'accompagne également d'un risque de surinterprétation de ses comportements. Il est possible que le participant cherche à attribuer une signification à un comportement anodin afin de satisfaire le chercheur qui organise l'entretien.

PERSPECTIVES

L'étude présentée dans les pages qui précèdent ne détaille qu'une seule expérimentation réalisée avec l'aide d'Alveho. Plusieurs itérations seront encore nécessaires pour affiner les observations qui en découlent. Ces itérations futures doivent permettre d'élargir l'échantillon de participants, de corriger des problèmes d'utilisabilité qui peuvent parasiter les résultats et d'éliminer progressivement les pistes infructueuses. Bien qu'il ne nous ait pas été permis de réaliser ces itérations dans le cadre de cette thèse, nous profitons de la présente section pour explorer des perspectives ressortant de nos observations et définir des pistes de recherche qui nous semblent prometteuses.

Correction des problèmes d'utilisabilité

Face aux difficultés d'utilisation rencontrées par nos participants, il nous faut d'abord distinguer celles qui relèvent de problèmes d'utilisabilité de celles qui relèvent plutôt d'une difficulté inhérente à la fonctionnalité et pour lesquelles il n'existe pas encore de « bonnes pratiques » de design. À titre d'exemple, la difficulté de manipulation de la page « Profile » par l'ensemble de nos participants pourrait être interprétée comme une difficulté inhérente à la compréhension du concept de « trace numérique » et à la façon dont des algorithmes s'adaptent de façon dynamique à l'évolution de ces traces. Toutefois, il est surprenant que tous nos participants partagent ces difficultés. Dans ce cas-ci, les réponses aux entretiens semblent indiquer que les participants de l'étude comprennent, de façon plus ou moins précise, que leur comportement est suivi par le site et que leurs lectures influencent l'algorithme, ce qui contredit l'hypothèse que nous venons de proposer. Si nous prenons en compte notre observation ultérieure sur le poids des habitudes de navigation ainsi que la littérature existante (Harambam *et al.*, 2019), il est fort probable que les difficultés rencontrées s'expliquent plutôt par des problèmes d'utilisabilité. Cette explication est également en phase avec nos hésitations lors de la phase de développement quant à la forme finale de ces composants.

Dans un cas comme dans l'autre, certaines problématiques sont plus faciles à catégoriser. Si nous avons un élément textuel mal placé qui est illisible parce qu'il se mélange à un autre élément visuel, le problème est plutôt du registre de l'utilisabilité. À l'inverse, les difficultés rencontrées par nos participants quant à la manipulation des poids de l'algorithme nous semblent traduire une difficulté inhérente à la représentation et à la compréhension du fonctionnement de l'algorithme et des actions qu'il est possible d'opérer sur lui. Ces deux catégories ne sont pas deux blocs hermétiques mais plutôt une distinction analytique qui doit surtout nous servir à distinguer ce qui relève de résultats pertinents qu'il est nécessaire d'explorer et ce qui relève plutôt d'une simple erreur de conception pour laquelle des solutions existent déjà.

À l'heure actuelle, il nous semble que les changements suivants peuvent nous aider à résoudre la majorité des problèmes d'utilisabilité identifiés lors de notre étude :

- Utiliser des types de graphiques distincts pour représenter des phénomènes différents ;
- Rapatrier les fonctionnalités de la page de profil sur la page d'accueil de sorte à regrouper les paramètres de contrôle en un seul lieu ;
- Ajouter des feedbacks visuels pendant le rechargement du fil d'actualité.

Le reste de cette section est consacré à l'exploration de perspectives nouvelles ouvertes par les résultats de cette étude. Nous explorerons notamment plusieurs pistes de design qui nous semblent prometteuses pour résoudre certaines difficultés rencontrées qui ne peuvent être réduites à une simple erreur de conception.

D'une éthique à une esthétique

Un premier angle mort de nos résultats concerne le passage d'une autonomie locale à une autonomie globale. L'autonomie globale d'un individu n'est pas seulement l'accumulation d'un nombre suffisant d'épisodes d'autonomie locale (Oshana, 2016). Même si elle préfère parler d'autonomie épisodique et d'autonomie programmatique, Diana T. Meyers (1989) offre l'explication suivante pour distinguer les deux notions :

Autonomous episodic self-direction occurs when a person confronts a situation, asks what he or she can do with respect to it - the options may include withdrawing from it as well as participating in it in various ways - and what he or she really wants to do with respect to it, and then executes the decision this deliberation yields. Autonomous

programmatic self-direction has a broad sweep. Instead of posing the question « What do I really want to do now? » this form of autonomy addresses a question like « How do I really want to live my life? » To answer this latter question, people must consider what qualities they want to have, what sorts of interpersonal relations they want to be involved in, what talents they want to develop, what interests they want to pursue, what goals they want to achieve, and so forth. Their decisions about these matters together with their ideas about how to effect these results add up to a life plan. (pp.48-49)

Lorsque nos participants parviennent à élaborer une réflexion critique, qu'elle soit plus ou moins approfondie, celle-ci se limite souvent au domaine d'Alveho et à son algorithme. Lorsque nous les avons interrogés sur leurs pratiques informationnelles pendant l'étude, nous avons pu constater qu'elles ne furent peu ou pas impactées par l'usage d'Alveho. Au mieux, certains participants ont arrêté de suivre l'actualité sur un autre site d'information pour le remplacer par Alveho dans le cadre de l'étude. Toutefois, nous n'avons jamais pu constater, lors de nos entretiens, le développement d'une réflexion qui porte plus largement sur leurs pratiques médiatiques globales. Ce constat est en partie lié aux limites de notre méthodologie qui ont contribué, pour certains de nos participants, à faire de la participation à l'étude une routine un peu ennuyante (à ce sujet, nous renvoyons à la section consacrée aux limites de notre méthodologie).

Nous espérons toutefois stimuler une réflexion qui porte, de façon plus fondamentale, sur leur rapport à la norme et à sur la forme que devraient prendre, selon eux, leurs habitudes de navigation et leur rapport à l'information. Dans son étude sur les pratiques d'écoute musicale sur LastFM, Karakayali *et al.* (2018) se sont intéressés à la plateforme d'écoute musicale *LastFM*. L'utilisateur peut connecter un *scrobbler* à différentes applications de son choix qui enregistre les pistes musicales écoutées et qui permet à *Lastfm* de composer une librairie personnelle sur base de ce qu'il écoute. Cette librairie alimentera la production de recommandations personnalisées. Les auteurs de l'étude constatent que cette gestion consciente de son historique de navigation ainsi que l'usage concomitant de la recommandation algorithmique contribuent à faire de leur profil musical un objet de préoccupation qu'il devient important de façonner. Chaque set de recommandations est interprété comme un reflet déformé de leur profil numérique qui indique une actualisation possible de leur identité musicale. Cet univers des possibles encourage les participants à identifier un

impératif qu'ils jugent important à suivre, ici la diversité de leur écoute, et à maintenir volontairement l'ambiguïté sur leur identité musicale :

We shall therefore suggest that the pursuit of diversity, accompanied by the constant flow of recommendations, calls for a virtually endless care and cultivation of «music taste». In fact, some users perceive «indeterminacy» as the trademark of their «music taste». (Karakayali *et al.*, 2018, p. 13)

En ce qui nous concerne, nous étions à la recherche d'une réflexion mettant en tension les normes imposées par l'algorithme (notamment l'identification de contenus similaires à leurs goûts) et une série de normes communément partagées au sein de la société (suivre une actualité diversifiée, se tenir au courant de l'actualité politique, privilégier certaines sources reconnues, garder un intérêt pour de nouvelles choses, etc.). À la manière de ce que Foucault qualifie d'une éthique de la liberté, nous espérions observer l'élaboration d'un positionnement par rapport aux contradictions et synergies potentielles pouvant être visibilisées par les outils mis à disposition. Le seul moment où nous avons observé un phénomène similaire fut lorsque certains participants de l'étude se sont interrogés sur les statistiques mises à disposition sur leurs pages de profil. Si la présence de statistiques apparaît chez Karakayali *et al.* (2018) comme un vecteur de stylisation de son identité musicale, ce genre de fonctionnalité peut également conduire à une forme de moralisation de la pratique de l'utilisateur et une incitation à la performance. C'est visiblement le deuxième scénario qui s'est réalisé dans notre cas. En effet, à cause d'un écart entre le nombre d'articles qu'ils ont lu et la moyenne de l'ensemble des participants qui était plus élevée, certains de nos participants se sont brièvement mis à douter de leurs pratiques. Dès lors, une piste de recherche potentielle concerne l'évaluation des facteurs, qu'il s'agisse de conditions de navigation ou de caractéristiques personnelles de l'utilisateur, qui sont plus à même d'orienter l'usage de ces informations du côté de la performance ou du côté d'une pratique créative.

Du souci de soi au souci des autres

On pourrait reprocher au souci de soi foucauldien ainsi qu'à notre conceptualisation de l'autonomie de privilégier une conception individualiste du sujet. Toutefois, comme le souligne très bien Gros (2002),

(...) pour Foucault, mais une fois encore dans le cadre historique qu'il décrit, celui des deux premiers siècles de notre ère, que le souci de soi tel qu'il est prôné par les moralistes, encouragé par les philosophes, prescrit par les sages stoïciens, ce souci de soi ne consiste pas à couper le sujet du monde, à le détacher des autres, à l'isoler, mais seulement à lui trouver *sa juste place* et à *leur* trouver, aux autres, leur juste place. (p.234)

Dès lors, l'autonomie que nous avons formalisée dans cette thèse se veut doublement sociale. Il s'agit à la fois d'une pratique qui s'exerce avec les autres, dans le cadre du travail d'introspection et d'élaboration de son identité, mais également d'un exercice qui vise à se situer dans un espace social particulier. À travers cette forme que nous donnons à notre existence, quelle est la place que nous nous donnons en tant qu'individu dans la collectivité qui est la nôtre et quelles sont les limites que nous nous imposons ?

Il nous faut toutefois reconnaître que la version actuelle d'Alveho ne permet pas ce passage du soi à la collectivité. Deux limitations techniques sont en partie responsables de ce constat. Pour commencer, les recommandations produites par l'algorithme d'Alveho sont basées sur une analyse textuelle des contenus des articles. Or, la plupart des recommandations que l'on retrouve sur des services commerciaux capitalisent également sur des informations de nature sociale relatives à leurs communautés d'utilisateurs. Si tel contenu m'est recommandé, ce n'est pas uniquement parce qu'il ressemble à des contenus que j'ai pu apprécier dans le passé mais également parce qu'il s'agit d'un objet qui a attiré l'attention d'une personne au profil similaire au mien. La popularité d'un contenu au sein de la communauté est un autre type d'indicateur qui peut également être mobilisé pour justifier une recommandation. Aucun de ces signaux n'est mesuré ou pris en compte par l'algorithme étant donné la durée de l'étude qui n'était pas suffisante pour nous permettre de récupérer assez de données, tandis qu'une recommandation basée sur le contenu pouvait être mise en place immédiatement. Étant donné que nous avons opté pour une méthodologie expérimentale avec une plateforme « vierge », nous ne disposons pas de données suffisantes sur la communauté d'utilisateurs, de toute façon trop petite, pour pouvoir offrir des recommandations collaboratives. De façon plus générale, il est plus difficile d'associer des techniques de recommandation collaboratives avec la recommandation d'actualités car la nature plus éphémère de ces contenus ne laisse pas assez de temps pour accumuler des données pertinentes (en tout cas, lorsque la

communauté d'utilisateurs est petite). Le contenu est déjà « périmé » lorsqu'il devient possible d'identifier des tendances communes entre des utilisateurs.

Une autre limitation concerne les conditions de réception de ces recommandations. La plupart des systèmes de recommandation contemporains s'insèrent dans un réseau social ou sur une plateforme qui permet une forme d'échanges plus ou moins approfondie avec la communauté d'utilisateurs. La présence de commentaires, de « likes », de partages ou encore de listes d'écoutes dans le cadre de plateformes musicales contribuent à faire de la réception d'une recommandation une expérience collective. Or, ces marqueurs semblent jouer un rôle crucial dans le processus de sélection et d'interprétation de l'information (Messing et Westwood, 2014). En printemps 2022, Darren Huang, un étudiant de l'école d'ingénieurs de l'UCLA développe Spotify Pie, un outil en ligne, basé sur l'API de Spotify, permettant à quiconque de générer un graphique circulaire représentant son historique d'écoute musicale sur base de ses données personnelles²⁸. D'une façon très similaire à ce que nous avons fait avec Alveho, ce graphique permet de visualiser la répartition des différents genres musicaux écoutés par l'utilisateur. L'outil a connu une grande popularité sur les réseaux sociaux à cette période. Cette popularité ne s'explique pas par son originalité ou sa complexité. En réalité, Spotify Pie repose sur des requêtes très simples à l'API de Spotify. Ce n'est pas non plus pour la richesse de l'information qui est extraite étant donné qu'il s'agit d'une représentation graphique tout à fait classique. S'il a eu autant de succès, c'est probablement pour le format des résultats produits par le dispositif qui permet à chaque utilisateur de partager la « forme » que prend son profil sur les réseaux sociaux.

Au-delà de l'implémentation d'un algorithme collaboratif, il nous semble surtout intéressant d'identifier l'influence de marqueurs sociaux sur l'appropriation de visualisation de données et de statistiques centrées sur la pratique de l'utilisateur. Peut-être pouvons-nous y trouver une solution au dilemme identifié dans la section précédente entre performance et esthétique. Pour finir, d'un point de vue méthodologique, complexifier le système de la sorte peut également permettre de rapprocher les conditions expérimentales à des conditions réelles de navigation.

²⁸ <https://huangdarren1106.github.io/#>

Leviers d'autolégitimation

Outre la nature du jugement réflexif développé par les participants de notre étude, notre analyse s'est également portée sur les facteurs influençant l'appropriation des outils implémentés. Si la compétence d'autonomie s'acquiert par la pratique, mais qu'initier cette pratique demande de s'auto-légitimer dans son action et de se reconnaître un niveau de compétence suffisant, comment faire en sorte qu'un individu non autonome puisse le devenir ? Au-delà d'un travail d'éducation aux médias, serait-il alors possible d'identifier des mécanismes de design qui puissent encourager l'utilisateur à voir dans ses pratiques informationnelles un sujet de préoccupation et, surtout, de se voir comme un sujet compétent à agir ? Si nous n'avons pas de réponse immédiate à ces questions, nous pouvons toutefois explorer deux modalités d'interaction homme-machine qui peuvent, selon nous, servir de tremplin à cette attitude.

Intrusion et interruption

Le design d'Alveho repose sur un principe général de non-intervention. L'ensemble des fonctionnalités implémentées servent soit à compenser l'asymétrie informationnelle entre l'utilisateur et le système ou à lui donner des moyens d'agir sur le fonctionnement de l'algorithme. Il est toutefois possible d'envisager des mécanismes intrusifs qui viennent interrompre le comportement de l'utilisateur.

Twitter s'est notamment intéressé à cette thématique à travers plusieurs expérimentations sur sa propre plateforme. La première qu'il est possible d'identifier concerne le partage d'articles. Afin de limiter la vitesse de diffusion d'informations fausses, l'entreprise expérimente en 2020 une fonctionnalité provoquant l'apparition d'un message préventif lorsqu'un utilisateur souhaite partager un tweet contenant un lien que l'utilisateur n'a lui-même pas ouvert. L'utilisateur doit alors cliquer sur un bouton de confirmation avant de pouvoir partager le tweet en question. Les détails de l'étude ne furent pas publiés et seuls quelques chiffres furent partagés par l'entreprise sur son propre site. Les utilisateurs seraient ainsi 40 % plus enclins à lire l'article après avoir lu le message. Une autre étude dont les résultats furent publiés en 2022 s'est intéressée à l'ajout d'un message d'avertissement avant la création d'un nouveau tweet afin de limiter la production de contenus haineux (Katsaros *et al.*, 2022) :

In this experiment, we designed an intervention aimed at interrupting those caught in a « hot state » while posting content online. This intervention would allow users to pause

and think about their comments to avoid regretful postings. As discussed more below, we hypothesize that intervening during this moment will not only have an impact at the time of intervention (H1) but can help change people's future behaviors (H2 and H3), and shape the conversations these individuals are participating in (H4). (p.479)



Figure 24 — Capture d'écran du message d'avertissement, issu de Katsaros, Yang et Fratamico (2022)

Nous ne ferons pas ici une critique approfondie de leur analyse. Selon leurs résultats, le groupe test a produit 6 % de messages contenant des propos offensants en moins que le groupe de contrôle. Si ces mécanismes sont probablement efficaces pour limiter des comportements impulsifs, il reste encore à prouver qu'ils puissent être efficaces pour stimuler des démarches réflexives plus ou moins durables (Bovens, 2009). Les auteurs de l'étude reconnaissent eux-mêmes que le succès modeste de cette intervention indique bien qu'il ne s'agit que d'une solution partielle à la problématique générale qu'elle est censée résoudre.

Toutefois, en complément d'autres mécanismes, une interruption mesurée de la navigation de l'utilisateur peut avoir un certain mérite lorsque celui-ci rentre dans une démarche que l'on pourrait qualifier de « mécanique » (comme faire défiler son fil d'actualité sans y porter réellement attention). Nous pouvons également faire l'hypothèse que cela peut encourager un utilisateur figé sur une configuration particulière à explorer de nouvelles modalités d'interaction. Un désaxage occasionnel pourrait même contribuer à rendre une certaine fluidité à une identité qui se serait figée dans des pratiques routinières, à l'image des utilisateurs de LastFM identifiés par Karakayali *et al.* (2018) qui maintiennent une incertitude autour de leurs goûts musicaux.

Nous avons déjà proposé d'imposer une limite dans le nombre de recommandations soumises. Une autre option pourrait consister à changer périodiquement la configuration de l'utilisateur tout en le prévenant de ces modifications. Il reste néanmoins à déterminer quand cette intrusion devient une nuisance à l'égard de l'expérience utilisateur et devient préjudiciable pour le développement d'une réflexion critique.

Anthropomorphisation de l'algorithme

L'évolution technique de certaines technologies d'intelligence artificielle encourage de nombreux utilisateurs à les anthropomorphiser, c'est-à-dire leur attribuer des caractéristiques humaines. Cette habitude est d'autant plus facile à adopter quand le dispositif a été conçu dès le départ avec des caractéristiques qui visent à rappeler le comportement d'un être humain. Les « assistants personnels » de types Amazon Echo ou Alexa sont les exemples les plus emblématiques de cette tendance. L'usage de la commande vocale, le qualificatif d'assistant personnel ou encore l'attribution d'un nom pouvant rappeler une personne humaine contribue à renforcer l'illusion.

L'anthropomorphisme a déjà été étudié dans le domaine de la robotique sociale (Breazeal, 2002) et des interactions humain-machine (Suchman, 2007). Plusieurs recherches ont déjà évalué son effet sur la confiance à l'égard d'un dispositif (Waytz *et al.*, 2014) ou encore de la facilité d'utilisation (Mori, 1970). Dans ces cas-là, l'objet étudié est généralement un robot et non une simple interface numérique. Quel que soit le domaine d'analyse, l'anthropomorphisation par un individu d'un artefact technique est interprétée comme une incapacité à appréhender la complexité technique de l'objet, réduisant par conséquent sa capacité à en évaluer les effets sur soi. Selon B. Friedman et Kahn (1992), il faut absolument empêcher l'utilisateur de projeter une agentivité humaine sur le dispositif technique car l'utilisateur risque d'inférer des capacités dont il ne dispose

pas. Autant Shneiderman (1989) que Winograd et Flores (1986) préconisent une interface la plus transparente possible pour permettre à l'utilisateur de se concentrer sur la tâche qu'il essaie de réaliser.

When a transparent interaction is achieved, the user is freed from the details of using the tool to focus on the task at hand. The shared vision here is for the interface to «disappear», not to intercede in the guise of another «agent» between human users and the computational system (B. Friedman et Kahn, 1992, pp.10-11)

Cette démarche qui fait tout son sens pour une multitude de contextes opérationnels contribue toutefois à faire de ces systèmes des boîtes noires et à invisibiliser l'algorithme (Eslami *et al.*, 2015). Dans le cadre du projet de design qui est le nôtre, c'est justement l'effet inverse que nous essayons de produire. Nous souhaitons que l'algorithme soit envisagé comme un agent intermédiaire dont le « comportement » est la traduction technique des intentions de son concepteur qui ne sont pas nécessairement alignées sur celles de l'utilisateur. Or, nous avons pu constater, dans la section précédente, les limites d'une interface qui se veut plus « objective » ou plus « rationnelle ». Nous proposons ici d'envisager une hypothèse alternative en envisageant l'anthropomorphisation comme un levier potentiel au développement d'une réflexion critique et, in fine, d'un usage autonome. Pour soutenir cette hypothèse, nous proposons d'envisager l'anthropomorphisation comme une métaphore conceptuelle (Lakoff et Johnson, 2003). Ce processus cognitif implique de projeter des dimensions d'un domaine connu sur un domaine plus abstrait et complexe afin d'en faire sens plus facilement. Dans ce cas-ci, en même temps que l'utilisateur projette certaines dimensions physiologiques de l'interaction entre êtres humains (la voix, l'apparence physique, etc.), il pourrait également projeter des dimensions plus abstraites liées à la confiance, le mensonge, la possession d'intentions (bonnes ou mauvaises). Ces caractéristiques pourraient stimuler chez l'utilisateur le développement d'une réflexion critique sur la « personnalité » de cet algorithme et de ses intentions, éléments qui sont habituellement camouflés par un imaginaire de l'outil neutre. Marenko et van Allen (2016) suggèrent notamment d'indiquer les biais d'outils de traitement de données en leur attribuant des personnalités qui pourraient stimuler la créativité de leur utilisateur.

Pour enclencher ce processus d'anthropomorphisation avec un système de recommandation, Harambam *et al.* (2018) suggèrent d'attribuer des personae à plusieurs algorithmes que l'utilisateur

peut sélectionner. Ces personae prendraient la forme de personnages fictifs dont la dénomination indiquerait métaphoriquement les biais et tendances de l'algorithme. Par exemple, un algorithme qui a pour attribut de proposer des contenus qui divergent des intérêts de l'utilisateur pourrait être qualifié d'« explorateur ». Un système configuré pour proposer des contenus issus de plusieurs bords politiques de façon équilibrée pourrait être nommé le « diplomate ». L'interface du système pourrait également préciser la valeur des paramètres de l'algorithme en dessous de l'intitulé de sorte que l'utilisateur puisse, à plus long terme, identifier le lien entre la dénomination de l'algorithme et le fonctionnement du système. À l'heure actuelle, nous n'avons pas pu identifier de travaux empiriques qui corroborent cette hypothèse.

Perspectives en éducation aux médias

Cadrage

L'ensemble du cadrage théorique et méthodologique présenté dans cette thèse fut mobilisé dans le cadre d'une pratique critique de design. Toutefois, l'ensemble de cet appareillage, autant théorique que méthodologique, peut également s'inscrire dans une démarche d'éducation. Pour conclure ce chapitre dédié aux résultats et aux perspectives issues du développement d'Alveho, nous présentons ici de façon succincte une activité d'éducation aux médias reposants sur la réflexion générale qui fut développée dans le cadre de cette thèse.

Outre les démarches éducatives consacrées au développement d'une connaissance opérationnelle, la majorité des démarches d'éducation critiques aux algorithmes s'inspirent des travaux de Freire ou De Certeau et reposent sur le développement d'une attitude de résistance de la part de l'utilisateur. Si cette approche a de nombreux mérites, elle a pour inconvénient de se limiter à l'élaboration d'une critique détachée des pratiques réelles des participants. En complément de ces approches reposant sur une identification des rapports de pouvoir à l'œuvre, nous proposons un prototype d'activité d'éducation aux médias qui encourage le développement d'un usage autonome. Bien que la capacité d'action des usagers d'outils numériques soit très limitée à l'heure actuelle, il reste malgré tout possible de proposer une éducation critique aux algorithmes qui repose sur un travail d'introspection et de négociation des relations de pouvoir qui s'établissent entre l'usager et le système technique.

La conception de ce dispositif découle d'une demande d'un organisme d'éducation aux médias (Média Animation). Étant donné la nature concrète de la demande et l'implication de praticiens,

notre approche s'apparente à celle de la recherche design en éducation (RDE) (Class et Schneider, 2013). La RDE se trouve à mi-chemin entre la recherche-action et la science du design en empruntant à l'un sa volonté transformative et à l'autre sa volonté de créer un savoir durable de design. La RDE est une pratique pragmatique qui s'insère au cœur de problématiques concrètes de praticiens. À travers une dynamique collaborative et itérative, elle vise à apporter un changement durable dans une situation éducative spécifique. Selon Class et Schneider (2013), elle s'organise autour d'une démarche cyclique qui se décompose en plusieurs étapes qui impliquent généralement les partenaires praticiens :

- L'évaluation d'un problème;
- La création d'un design provisoire sur base de la littérature ;
- L'évaluation du prototype dans un objectif d'affinement de l'artefact ainsi que des hypothèses théoriques à son origine.

Cette méthodologie a pour finalité scientifique la formalisation de règles de design en éducation. À l'inverse d'un protocole expérimental plus rigide qui vise à éprouver une série d'hypothèses, cette méthodologie vise à prendre en compte toute la complexité d'une situation d'éducation réelle, en ce compris les considérations opératives. Le premier cycle de cette démarche a été présenté dans le troisième chapitre. Des professionnels de l'éducation aux médias issus de l'organisme à l'origine de notre démarche furent invités à plusieurs reprises à donner leur avis sur les développements en cours.

Réalisation

Après la finalisation de la première version d'Alveho qui a servi à réaliser notre expérimentation en navigation libre, nous avons mobilisé la plateforme dans le cadre d'une première activité d'éducation aux médias (EAM). Celle-ci s'est déroulée avec 21 participants issus de l'ASBL d'éducation aux médias Action Médias Jeunes (ACMJ). Ils étaient donc familiers avec les enjeux de cette activité. L'atelier était divisé en deux étapes : l'activité d'EAM à proprement parler, suivie d'un moment de retour de la part des participants sur le déroulement de l'activité et sur Alveho. L'activité d'EAM était, quant à elle, divisée en 5 étapes:

1. **Initiation** : Découverte libre de la plateforme avec pour objectif de comprendre comme fonctionne l'algorithme ;

2. **Théorie** : Présentation d'éléments techniques généraux sur les algorithmes de recommandation ;
3. **Comparaison** : Attribution d'un profil préconfiguré par groupe. Chaque groupe devait ensuite résumer son fil d'actualité et comparer ses résultats avec les autres groupes ;
4. **Manipulation** : Chaque groupe avait pour objectif de modifier son fil d'actualité pour atteindre un objectif ciblé en utilisant tous les outils à sa disposition (ex : s'ils avaient beaucoup de recommandations sportives, ils devaient essayer d'obtenir plus de recommandations d'articles culturels) ;
5. **Ouverture** : L'activité se termine sur un débat ouvert avec tous les participants pour introduire une réflexion plus abstraite sur les enjeux sociaux relatifs à ces technologies (libre arbitre, découverte de contenus etc.).

Étant donné les mesures sanitaires en place à cause de la pandémie de covid lors de la réalisation de cette activité, celle-ci a dû être réalisée à distance en visioconférence. Pour permettre aux participants de chaque groupe de se concerter et d'agir ensemble sur Alveho, un modérateur était désigné par groupe pour partager son écran et ses actions sur le site.

À travers cette première itération, plusieurs participants ont signalé que manipuler l'algorithme leur a permis de comprendre des notions abstraites qu'ils n'arrivaient pas à appréhender avant cette activité. Toutefois, le résultat le plus saillant concerne la dimension collective de la manipulation. Demander aux participants de manipuler ensemble le système pour ensuite comparer leurs résultats avec les autres groupes a conduit à stimuler les échanges au cours de l'activité, ce qui a permis aux modérateurs de rapidement embrayer sur un discours critique relatif à certains enjeux sociaux plus abstraits et à mettre ces enjeux en perspective avec les pratiques des participants. La réalisation de cette activité a également permis de confirmer certains problèmes d'utilisabilité identifiés lors de l'étude principale.

À la suite de cette première itération, nous avons entamé plusieurs allers-retours avec nos partenaires praticiens pour affiner à la fois l'outil et l'activité. Plusieurs itérations seront encore nécessaires pour finaliser cette activité. Cette expérimentation préliminaire nous permet toutefois de souligner l'intérêt de disposer d'un outil à manipuler pour ancrer les échanges critiques, souvent abstraits, dans un exercice concret d'éducation aux médias.

CONCLUSION

Un système de recommandation est un dispositif informatique qui produit de façon automatique des recommandations en fonction des informations dont il dispose sur l'utilisateur du service. L'intégration de ces dispositifs au sein d'un service en ligne permet d'offrir à l'utilisateur une porte d'accès personnalisée à un catalogue d'objets trop important pour être consulté dans sa totalité. En réalisant l'état des lieux des enjeux sociaux qui accompagnent l'usage de ces dispositifs, nous avons très rapidement fait le constat que la forme actuelle de ces systèmes limite la capacité de leurs usagers à développer un regard critique sur leurs effets. Le recours à un design qui se veut « lisse » et « transparent » pour l'utilisateur conduit à un effacement de l'algorithme du premier plan de l'interface pour opérer en tâche de fond. S'il reste possible pour un utilisateur expérimenté d'appréhender son fonctionnement, il devient de plus en plus difficile pour la majorité des usagers d'évaluer son effet sur l'ordonnancement de l'information. Cette asymétrie informationnelle, entre d'une part ce que le système connaît de l'utilisateur et d'autre part ce que l'utilisateur connaît du système, conduit à un déséquilibre dans les relations de pouvoir qui s'établissent entre les usagers et les concepteurs de ces dispositifs. Toujours dans l'optique de faciliter l'expérience utilisateur, ces systèmes sont généralement dépourvus de mécanismes permettant d'altérer leur configuration et d'agir directement sur leur comportement. L'automatisation complète du processus de recommandation vise ainsi à limiter l'implication de l'utilisateur de sorte à lui permettre de consacrer son attention sur les contenus qui lui sont recommandés. Si cette pratique conduit à une optimisation de l'expérience utilisateur, elle prive progressivement l'utilisateur de toute prise sur laquelle il pourra s'appuyer pour construire une attitude critique relative à ses usages.

Ce constat nous a amené à mettre au point un projet de recherche ayant pour objectif d'identifier une trajectoire de conception alternative qui puisse réincorporer de la friction dans la relation entre l'utilisateur et le système. Sans pour autant renoncer à une expérience utilisateur fluide et plaisante, nous avons cherché à réintroduire des mécanismes permettant à l'utilisateur d'appréhender la complexité du couplage entre ses pratiques et le système automatique de recommandation de sorte à soutenir le développement d'une forme de réflexivité critique. Dit autrement, nous ne cherchons pas à expliquer à l'utilisateur comment être lisible pour le système. Au contraire, nous souhaitons mettre en évidence l'effet du dispositif sur ses pratiques tout en fournissant à l'utilisateur les outils nécessaires pour qu'il puisse moduler la relation qu'il établit avec ce dispositif.

Afin de formaliser cette idée d'une réflexivité critique, nous avons décidé de recourir à la notion d'autonomie. Notre intérêt pour ce concept s'explique notamment pour son usage fréquent, dans une littérature multidisciplinaire, en contrepoint de la notion de bulle de filtres. Nous avons souligné la nature problématique de cette notion qui tend à simplifier à plusieurs niveaux la relation entre l'utilisateur et le système technique. Toutefois, les réflexions de design qui se sont cristallisées en réponse à cette notion conservent leur pertinence et mettent en évidence une préoccupation commune pour une implication plus active de l'utilisateur dans le processus de recommandation.

Derrière cette notion d'autonomie se cache une richesse sémantique qu'il ne fut pas possible d'explorer en profondeur dans le cadre de cette thèse. L'interprétation principale sur laquelle nous nous sommes arrêtés se retrouve dans le domaine de l'interaction humain-machine autour de la notion de « réalisation de soi », elle-même inspirée de la psychologie humaniste. Selon ce paradigme, un individu acquiert une autonomie individuelle lorsqu'il parvient à identifier les intentions réelles d'un « soi profond ». Ce soi profond, qui préexiste à toute forme d'influence extérieure, incarnerait un potentiel inné que chaque individu est libre d'actualiser à condition de se retrouver dans des conditions favorables. De la même façon que la psychologie humaniste s'est construite en opposition au paradigme behavioriste dominant de l'époque, l'intégration de cette conceptualisation de l'autonomie dans le domaine de l'interaction humain-machine s'est faite en opposition à une approche de design reposant sur une analyse comportementaliste de l'utilisateur. Les principales perspectives de design issues de cette approche (Knijnenburg *et al.*, 2016) visent à permettre l'identification de besoins qui seraient préexistants à toute forme de médiation technique.

Si cette approche a eu le mérite de donner naissance à plusieurs prototypes originaux, elle est toutefois sensible aux mêmes critiques que celles qui furent formulées à l'égard de la psychologie humaniste. En postulant un soi profond qui serait ontologiquement distinct de toute forme de socialité ou de technicité, cette approche rentre en contradiction avec les théories les plus contemporaines issues des sciences cognitives et de la philosophie des techniques sur la nature de notre rapport aux objets. Le paradigme de la cognition distribuée, issu des sciences cognitives, a ainsi contribué à mettre en évidence que nos capacités cognitives ne sont pas simplement étendues par nos artefacts cognitifs, mais que ces derniers contribuent à l'émergence de processus cognitifs qui ne seraient pas possibles en leur absence. Au-delà d'opérations cognitives de base, notre usage d'artefacts techniques agit également sur des niveaux de représentations plus complexes relatifs à

notre identité morale. Plusieurs travaux issus de la sociologie et de la philosophie des techniques ont contribué à mettre en évidence la façon dont notre environnement technique influe sur nos comportements et, plus fondamentalement, contribue à modeler nos conditions d'existence. Notre expérience de la conduite est activement modelée par l'infrastructure routière. De la même façon, l'élaboration de son identité en tant que parent ainsi que sa relation à son enfant commence dès la grossesse à travers l'imagerie médicale contemporaine. Ces perspectives nous invitent à concevoir notre rapport aux objets comme une relation de pouvoir bidirectionnelle plutôt que comme une relation unidirectionnelle selon laquelle un individu exercerait sa pleine maîtrise sur un objet neutre. Si nous agissons sur l'objet, celui-ci agit également sur nous en retour. Dès lors, depuis notre point de vue relationnel, cette interprétation de l'autonomie comme réalisation d'un soi profond n'est pas capable de rendre compte de la contingence entre les comportements de l'utilisateur et l'environnement technique dans lequel celui-ci se déploie.

Pour en revenir plus spécifiquement aux systèmes de recommandation, une littérature croissante focalisée sur les usages et les modes d'appropriation de ces dispositifs contribue à mettre en évidence les allers-retours complexes qui se réalisent au sein de cette relation entre l'utilisateur et l'algorithme. L'algorithme n'est pas un outil neutre aux mains de l'utilisateur, mais celui-ci n'est pas pour autant le pantin d'une structure dominante.

À la suite de ce constat, plutôt que de renoncer entièrement à la possibilité d'un individu autonome dans son rapport à la technique, nous avons cherché une voie intermédiaire qui nous permette de maintenir cette idée d'un agent autonome tout en respectant notre cadrage ontologique général. Pour ce faire, nous avons proposé de passer par Foucault et, notamment, par ses derniers travaux portant sur le sujet. À travers son analyse du souci de soi dans l'antiquité grecque et romaine, Foucault identifie un rapport à la morale qui n'est pas de l'ordre du respect du code, mais plutôt de l'élaboration de sa propre vie comme d'une œuvre d'art. Cette esthétique de l'existence se manifeste à travers l'exercice de techniques de soi qui permettent à l'individu de revenir vers lui-même pour interroger le rapport qu'il décide d'établir à l'égard d'un système moral qui lui est présenté. Dans ce contexte, la norme n'est pas un cadre strict, mais plutôt un repère autour duquel l'individu se positionne à travers différentes techniques de retour à soi.

Ce retour à l'antiquité ne traduit pas, chez Foucault, l'apologie d'une sagesse antique qu'il s'agirait de réactiver. Toutefois, il estime que ce mode de rapport à soi peut inspirer un nouveau rapport à la norme qui repose sur une dynamique de stylisation de son existence.

Le modèle de la psychologie humaniste formalise l'autonomie comme une libération de toute forme d'influence extérieure de sorte à retrouver un noyau primaire qui serait le soi profond de l'individu. A l'inverse, en présentant la liberté comme une pratique esthétique et éthique de soi sur soi, Foucault nous permet d'envisager l'autonomie comme un travail permanent de régulation du nœud de relations de pouvoirs qui s'entremêlent autour de l'individu. Être autonome n'est pas un refus d'influences extérieures, mais plutôt leur modération consciente dans le cadre d'une création esthétique de soi-même.

En complétant cette approche par les travaux en philosophie morale de Meyers (1989) et de Mackenzie (2014), nous en sommes venus à proposer une définition de l'autonomie comme une compétence créative distribuée qui repose sur une attitude critique d'auto-légitimation. Être autonome au regard de son usage de systèmes de recommandation n'implique donc pas de se protéger de l'influence de ces dispositifs en limitant catégoriquement leur utilisation. Être autonome demande plutôt de façonner son écosystème médiatique et technique de sorte à délimiter les influences que l'on accepte d'endosser car leurs effets sont jugés bénéfiques au regard du plan de vie que l'individu s'est donné. Dans le cadre d'activités médiatiques, cela demande donc à l'individu de reconnaître un sujet de préoccupation dans ses pratiques médiatiques, d'identifier les logiques de distribution de l'information qu'il souhaite préserver et d'agir de sorte à configurer son environnement informationnel autour de ces logiques de distribution. Cette formalisation de l'autonomie nous a permis d'identifier 3 impératifs dont l'implémentation au sein de l'interface d'un système de recommandation doit permettre de soutenir le développement d'une telle forme d'autonomie : permettre à l'utilisateur d'agir sur le système (agentivité), garantir que l'utilisateur se sente légitime d'agir sur le système (légitimité) et rendre intelligible les effets de la relation qui s'établit entre l'utilisateur et le système (intelligibilité).

Une multitude de modalités d'interaction humain-machine peuvent être envisagées pour incarner ces 3 impératifs. Dans le cadre de cette thèse, nous avons décidé d'évaluer, plus spécifiquement, l'effet de composants de contrôlabilité de l'algorithme sur le développement de l'autonomie de l'usager. Ces éléments doivent lui permettre d'ajuster les paramètres de l'algorithme de sorte à

générer de nouvelles recommandations à la volée sur base des nouvelles configurations. À travers le développement de notre prototype Alveho, nous avons implémenté cette contrôlabilité à trois étapes différentes du processus de recommandation : à l'entrée des données, au niveau du processus et à la sortie des recommandations. Ces mécanismes furent accompagnés de plusieurs outils de visualisation du fil d'actualité et de l'historique de navigation.

Pour évaluer ce prototype, nous avons décidé de l'éprouver en situation réelle. Les méthodologies plus traditionnelles issues du domaine de l'expérience utilisateur reposent généralement sur des mises en situation uniques dans le cadre d'expérimentations en laboratoire. En ce qui nous concerne, étant donné la nature du domaine d'application du prototype ainsi que le type de fonctionnalités que nous évaluons, nous avons opté pour une mise en situation plus proche de conditions réelles d'usages de sorte à obtenir des résultats plus pertinents. En effet, suivre l'actualité n'est pas une pratique qui implique d'ajuster le système de recommandation à chaque session de navigation. Nous avons plutôt estimé que plusieurs sessions de navigation seront nécessaires pour que l'utilisateur s'approprie le dispositif et configure progressivement les paramètres en fonction de ses besoins. C'est donc pour ces raisons que nous avons décidé de mettre notre prototype à disposition de 23 participants, depuis leurs ordinateurs personnels et ce pour une durée de 5 semaines. Ces 5 semaines de navigation furent entrecoupées, pour chaque participant, d'une session de navigation en laboratoire sur une machine équipée d'un oculomètre ainsi que d'un entretien final. La première partie de cet entretien comprenait des questions nous permettant de cerner le niveau d'aisance en informatique de nos participants, d'identifier leurs pratiques médiatiques habituelles et leurs connaissances relatives à la recommandation automatisée. La seconde partie de l'entretien était une visite guidée, réalisée par le participant, de son compte Alveho pendant laquelle il lui était demandé d'expliquer sa routine de navigation pendant la durée de l'étude. À travers cette méthode mixte, mêlant le suivi précis des interactions de l'utilisateur avec des entretiens qualitatifs, nous estimons qu'il est possible d'avoir une vue particulièrement détaillée sur les comportements de nos participants.

Nos résultats nous indiquent que la présence d'outils de contrôle n'est pas une condition suffisante à leur utilisation et au développement d'un usage autonome. En effet, dans l'ensemble, les participants de l'étude conservèrent une attitude passive lors de leurs sessions de navigation. Le volume des logs d'activités nous indique que les participants de l'étude se sont surtout contentés

de faire défiler leur fil d'actualité. Une première explication concerne la charge cognitive qu'impose l'interactivité du dispositif et qui a déstabilisé plusieurs de nos participants. Une seconde explication notable concerne le décalage entre leurs habitudes de navigation et ce qui leur était demandé. Il était ainsi plus facile pour nos participants de mobiliser des techniques de navigation issues de leur expérience avec d'autres plateformes que de développer de nouveaux comportements mieux adaptés aux nouveautés du prototype expérimental. Malgré cela, plusieurs fonctionnalités semblent soutenir le développement d'une forme d'autonomie. En l'occurrence, permettre la comparaison de ses recommandations avec d'autres utilisateurs, interrompre le flux de l'activité et présenter, sous certaines conditions, des statistiques de navigation semblent des leviers prometteurs pour le développement d'une autonomie individuelle.

Au-delà des technologies de recommandations étudiées dans le cadre de ce travail, les résultats que nous présentons permettent également d'éclairer d'autres contextes d'usages qui mobilisent d'autres technologies de prise de décision automatisée. En effet, les tendances actuelles de lissage de l'expérience utilisateur ne concernent pas que les systèmes de recommandation. Pourtant, vivre en bonne intelligence avec ces dispositifs implique également de penser le degré de friction nécessaire à un usage critique et prudent. Qu'elle prenne la forme d'une frustration, d'une gêne, d'une interruption, cette friction vise à interrompre les automatismes et à encourager la prise de recul. La thèse générale que nous avons présentée vise à interroger le degré de friction qu'il est possible de réintroduire dans ces interfaces sans pour autant déstabiliser les utilisateurs les moins aguerris. Si nos résultats ne nous permettent pas de répondre de façon tranchée à cette problématique, ils nous permettent néanmoins d'identifier plusieurs pistes prometteuses permettant d'éveiller les utilisateurs de ces dispositifs aux enjeux normatifs de ces technologies.

RÉFÉRENCES

- Abbas, Y. et Dervin, F. (dir.). (2009). *Digital technologies of the self*. Cambridge Scholars.
- Achterhuis, H. (1998). *De erfenis van de utopie*. Ambo.
- Achterhuis, H. (dir.). (2001). *American philosophy of technology: the empirical turn*. Indiana University Press.
- Aiello, L. M. et Barbieri, N. (2017). Evolution of Ego-networks in Social Media with Link Recommendations. Dans *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 111-120. <https://doi.org/10.1145/3018661.3018733>
- Álvarez, M. V., López, J. M. T. et Ruíz, M. J. U. (2020). What Are You Offering?: An Overview of VODs and Recommender Systems in European Public Service Media. Dans Á. Rocha, C. Ferrás, C. E. Montenegro Marin et V. H. Medina García (dir.), *Information Technology and Systems* (p. 725-732). Springer International Publishing. https://doi.org/10.1007/978-3-030-40690-5_69
- Amatriain, X. et Basilico, J. (2015). Recommender Systems in Industry: A Netflix Case Study. Dans F. Ricci, L. Rokach et B. Shapira (dir.), *Recommender Systems Handbook* (2^e éd.). Springer US. <https://doi.org/10.1007/978-1-4899-7637-6>
- Anderson, A., Maystre, L., Anderson, I., Mehrotra, R. et Lalmas, M. (2020). Algorithmic Effects on the Diversity of Consumption on Spotify. Dans *Proceedings of The Web Conference 2020*, 2155-2165. <https://doi.org/10.1145/3366423.3380281>
- Bakardjieva, M. et Gaden, G. (2012). Web 2.0 Technologies of the Self. *Philosophy & Technology*, 25(3), 399-413. <https://doi.org/10.1007/s13347-011-0032-9>
- Bernstein, A., Vreese, C., Helberger, N., Schulz, W., Zweig, K., Baden, C., Beam, M., Hauer, M., Heitz, L., Jürgens, P., Katzenbach, C., Kille, B., Klimkiewicz, B., Loosen, W., Moeller, J., Radanovic, G., Shani, G., Tintarev, N., Tolmeijer, S. et Zueger, T. (2021). Diversity in News Recommendation. *Dagstuhl Manifestos*, 9.
- Beuscart, J.-S., Coavoux, S. et Maillard, S. (2019). Les algorithmes de recommandation musicale et l'autonomie de l'auditeur. *Rezeaux*, 213(1), 17-47.
- Blankenship, J. et Dansereau, D. F. (2000). The Effect of Animated Node-Link Displays on Information Recall. *The Journal of Experimental Education*, 68(4), 293-308.
- Bonenfant, M. et Farmer, Y. (2012). Je suis libre moi non plus : Facebook et la production de soi. Dans O. Kane, C. Perraton et F. Dumais (dir.), *Mobilisation de l'objet technique dans la production de soi*. Presses de l'Université du Québec.
- Bostandjiev, S., O'Donovan, J. et Höllerer, T. (2012). TasteWeights: a visual interactive hybrid recommender system. Dans *Proceedings of the sixth ACM conference on Recommender systems*, 35-42. <https://doi.org/10.1145/2365952.2365964>

- Bovens, L. (2009). The Ethics of Nudge. Dans T. Grüne-Yanoff et S. O. Hansson (dir.), *Preference Change : Approaches from Philosophy, Economics and Psychology* (p. 207-219). Springer Netherlands. https://doi.org/10.1007/978-90-481-2593-7_10
- Boy, J., Rensink, R. A., Bertini, E. et Fekete, J.-D. (2014). A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1963-1972. <https://doi.org/10.1109/TVCG.2014.2346984>
- boyd, danah. (2010). Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. Dans Z. Papacharissi (dir.), *A Networked Self*. Routledge.
- Breazeal, C. L. (2002). *Designing sociable robots*. MIT Press.
- Brin, S. et Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1), 107-117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Bruns, A. (2019). *Are filter bubbles real?* Polity Press.
- Bucher, T. (2017). The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, 20(1), 30-44. <https://doi.org/10.1080/1369118X.2016.1154086>
- Bucher, T. (2018). *If...Then: Algorithmic Power and Politics*. Oxford University Press. <https://www.oxfordscholarship.com/view/10.1093/oso/9780190493028.001.0001/oso-9780190493028>
- Bucher, T. et Helmond, A. (2018). The Affordances of Social Media Platforms. Dans J. Burgess, A. Marwick et T. Poel (dir.), *The SAGE Handbook of Social Media* (p. 233-253). SAGE Publications Ltd. <https://doi.org/10.4135/9781473984066>
- Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331-370. <https://doi.org/10.1023/A:1021240730564>
- Bush, V. (1945, 9 octobre). As We May Think. *Life Magazine*, 112-124.
- Cardon, D. (2015). *À quoi rêvent les algorithmes ?* Seuil.
- Castells, P., Hurley, N. et Vargas, S. (2015). Novelty and Diversity in Recommender Systems. Dans *Recommender Systems Handbook* (2^e éd., p. 881-918). Springer. https://doi.org/10.1007/978-1-4899-7637-6_26
- Chandesris, M. (2016). La forme des décisions algorithmiques : quels enjeux pour le design ? *Sciences du Design*, 4(2), 72-82. <https://doi.org/10.3917/sdd.004.0072>
- Christman, J. (2004). Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves. *Philosophical Studies*, 117(1), 143-164. <https://doi.org/10.1023/B:PHIL.0000014532.56866.5c>
- Christman, J. (2009). *The Politics of Persons: Individual Autonomy and Socio-historical Selves*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511635571>

- Cinquième contrat de gestion consolidé de la RTBF incluant son premier avenant, (2020).
- Clark, A. et Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7-19. <https://www.jstor.org/stable/3328150>
- Class, B. et Schneider, D. (2013). La Recherche Design en Education: vers une nouvelle approche ? *Francite.net*, (13), 5-16.
- Claussen, J., Peukert, C. et Sen, A. (2019). *The Editor vs. the Algorithm: Returns to Data and Externalities in Online News* (cesifo working paper series n° 8012). CESifo. https://ideas.repec.org/p/ces/ceswps/_8012.html
- Conseil de l'Europe. (2007). *Recommendation of the Committee of Ministers to member states on the remit of public service media in the information society*. Conseil de l'Europe.
- Conseil Supérieur de l'Audiovisuel. (2022). *Bilan du contrat de gestion de la RTBF (2019-2022)* (p. 127). Conseil Supérieur de l'Audiovisuel.
- Datta, H., Knox, G. et Bronnenberg, B. J. (2017). Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery. *Marketing Science*, 37(1), 5-21. <https://doi.org/10.1287/mksc.2017.1051>
- Davis, J. L. (2020). *How artifacts afford: the power and politics of everyday things*. The MIT Press.
- Davis, J. L. et Chouinard, J. B. (2016). Theorizing Affordances: From Request to Refuse. *Bulletin of Science, Technology & Society*, 36(4), 241-248. <https://doi.org/10.1177/0270467617714944>
- de Vries, K. (2010). Identity, profiling algorithms and a world of ambient intelligence. *Ethics and Information Technology*, 12(1), 71-85. <https://doi.org/10.1007/s10676-009-9215-9>
- Debray, R. (1995). *Vie et mort de l'image : Une histoire du regard en Occident*. Folio.
- Denouël, J. (2012). De l'utilisateur au concepteur, et retour. Quelques éléments sur l'intégration de l'expérience des usagers dans les processus d'innovation. *Echappées - ESA Pyrénées-Pau Tarbes*, (1), 67-75.
- Drumond, G. S. M., Coutant, A. et Millerand, F. (2018). La production de l'utilisateur par les algorithmes de Netflix. *Les Enjeux de l'information et de la communication*, 192(2), 29-44. <https://www.cairn.info/revue-les-enjeux-de-l-information-et-de-la-communication-2018-2-page-29.htm>
- Dubois, E. et Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729-745. <https://doi.org/10.1080/1369118X.2018.1428656>
- Dufasne, M. (2011). Plus de débats sur l'Europe pour plus d'implication des citoyens ? Analyse du forum de discussion en ligne « Debate Europe ». Dans C. Cheneviere et G. Duchenne (dir.), *Les modes d'expression de la citoyenneté européenne : actes de la journée d'étude*

organisée à Louvain-la-Neuve le 19 mars 2010 (p. 103-119). Presses universitaires de Louvain.

Dworkin, G. (1976). Autonomy and Behavior Control. *The Hastings Center Report*, 6(1), 23-28. <https://doi.org/10.2307/3560358>

Dworkin, G. (2017). Autonomy. Dans *A Companion to Contemporary Political Philosophy* (p. 439-451). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781405177245.ch18>

Dylko, I., Dolgov, I., Hoffman, W., Eckhart, N., Molina, M. et Aaziz, O. (2017). The dark side of technology: An experimental investigation of the influence of customizability technology on online political selective exposure. *Computers in Human Behavior*, 73, 181-190. <https://doi.org/10.1016/j.chb.2017.03.031>

Eger, N., Ball, L. J., Stevens, R. et Dodd, J. (2007). Cueing retrospective verbal reports in usability testing through eye-movement replay. Dans T. C. Ormerod et C. Sas (dir.), *Proceedings of the 21st British HCI Group Annual Conference on HCI 2007: HCI...but not as we know it – Volume 1, BCS HCI 2007, University of Lancaster, United Kingdom, 3-7 September 2007* (p. 129-137). BCS. <https://dl.acm.org/citation.cfm?id=1531312>

Ekstrand, M. D., Kluver, D., Harper, F. M. et Konstan, J. A. (2015). Letting Users Choose Recommender Algorithms: An Experimental Study. Dans *Proceedings of the 9th ACM Conference on Recommender Systems*, 11-18. <https://doi.org/10.1145/2792838.2800195>

Ekstrand, M. D. et Willemsen, M. C. (2016). Behaviorism is Not Enough: Better Recommendations through Listening to Users. Dans *Proceedings of the 10th ACM Conference on Recommender Systems*, 221-224. <https://doi.org/10.1145/2959100.2959179>

Ertzscheid, O. (2009). L'homme, un document comme les autres. *Hermes, La Revue*, n° 53(1), 33-40.

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K. et Sandvig, C. (2015). « I Always Assumed That I Wasn'T Really That Close to [Her] » : Reasoning About Invisible Algorithms in News Feeds. Dans *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 153-162. <https://doi.org/10.1145/2702123.2702556>

Etalab. (2022). *Expliquer les algorithmes publics*.

Regulation 2022 of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, (2022).

Fastrez, P. et Philippette, T. (2017). Un modèle pour repenser l'éducation critique aux médias à l'ère du numérique. *Tic&société*, (Vol. 11, N° 1), 85-110. <https://doi.org/10.4000/ticetsociete.2266>

Ferster, C. B. et Skinner, B. F. (1957). *Schedules of reinforcement*. Appleton-Century-Crofts.

- Fields, B., Jones, R. et Cowlshaw, T. (2018). The Case for Public Service Recommender Algorithms. Dans *FATREC workshop on Responsible Recommendation*. <https://piret.gitlab.io/fatrec2018/program/fatrec2018-fields.pdf>
- Fletcher, R. et Nielsen, R. K. (2017). Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication. *Journal of Communication*, 67(4), 476-498. <https://doi.org/10.1111/jcom.12315>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. et Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Fogg, B. J. (2003). *Persuasive technology: using computers to change what we think and do*. Morgan Kaufmann Publishers.
- Foucault, M. (1994). Qui êtes-vous professeur Foucault ? Dans *Dits et écrits : 1954 - 1988*. (vol. 1, p.601-602). Gallimard.
- Foucault, M. (1997). *Le souci de soi*. Gallimard.
- Foucault, M. (1999a). A propos de la généalogie de l'éthique : un aperçu du travail en cours. Dans *Dits et écrits : 1954 - 1988*. (vol. 4, p. 609-630). Gallimard.
- Foucault, M. (1999). L'écriture de soi. Dans *Dits et écrits : 1954 - 1988*. (vol. 4, p. 415-430). Gallimard.
- Foucault, M. (1999b). L'éthique du souci de soi comme pratique de la liberté. Dans *Dits et écrits : 1954 - 1988*. (vol. 4, p. 708-729). Gallimard.
- Foucault, M. (1999c). L'herméneutique du sujet. Dans *Dits et écrits : 1954 - 1988*. (vol. 4, p. 353-365). Gallimard.
- Foucault, M. (1999d). La poussière et le nuage. Dans *Dits et écrits : 1954 - 1988*. (vol. 4, p. 10-19). Gallimard.
- Foucault, M. (1999e). Le souci de la vérité. Dans *Dits et écrits : 1954 - 1988*. (vol. 4, p. 636-645). Gallimard.
- Foucault, M. (1999f). Le sujet et le pouvoir. Dans *Dits et écrits : 1954 - 1988*. (vol. 4, p. 222-242). Gallimard.
- Foucault, M. (1999g). Les techniques de soi. Dans *Dits et écrits : 1954 - 1988*. (vol. 4, p. 783-812). Gallimard.
- Foucault, M. (1999h). Qu'est-ce que les Lumières ? Dans *Dits et écrits : 1954 - 1988*. (vol. 4, p. 562-577). Gallimard.
- Foucault, M. (1999i). Une esthétique de l'existence. Dans *Dits et écrits : 1954 - 1988*. (vol. 4, p. 730-734). Gallimard.

- Foucault, M. (1999j). Usage des plaisirs et techniques de soi. Dans *Dits et écrits : 1954 - 1988*. (vol. 4, p. 539-561). Gallimard.
- Foucault, M. (2013). *L'origine de l'herméneutique de soi : conférences prononcées à Dartmouth College, 1980* (H.-P. Fruchaud et D. Lorenzini, dir.). Vrin.
- Foucault, M. (2015). *Qu'est-ce que la critique ? Suivie de La culture de soi* (Vrin).
- Foucault, M., Fruchaud, H.-P. et Lorenzini, D. (2017). *Dire vrai sur soi-même : conférences prononcées à l'Université Victoria de Toronto, 1982*. Vrin.
- Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1), 5-20. <https://doi.org/10.2307/2024717>
- Frankfurt, H. G. (1999). *Necessity, volition, and love*. Cambridge University Press.
- Friedman, B. (1996). Value-sensitive design. *Interactions*, 3(6), 16-23. <https://doi.org/10.1145/242485.242493>
- Friedman, B. et Kahn, P. H. (1992). Human agency and responsible computing: Implications for computer system design. *Journal of Systems and Software*, 17(1), 7-14. [https://doi.org/10.1016/0164-1212\(92\)90075-U](https://doi.org/10.1016/0164-1212(92)90075-U)
- Friedman, M. (1989). Feminism and Modern Friendship: Dislocating the Community. *Ethics*, 99(2), 275-290. <https://www.jstor.org/stable/2381435>
- Friedman, M. A. (1986). Autonomy and the Split-Level Self. *The Southern Journal of Philosophy*, 24(1), 19-35. <https://doi.org/10.1111/j.2041-6962.1986.tb00434.x>
- Geller, L. (1982). The Failure of Self-Actualization Theory: A Critique of Carl Rogers and Abraham Maslow. *Journal of Humanistic Psychology*, 22(2), 56-73. <https://doi.org/10.1177/0022167882222004>
- Gibbs, B. (1979). Autonomy and Authority in Education. *Journal of Philosophy of Education*, 13(1), 119-132. <https://doi.org/10.1111/j.1467-9752.1979.tb00527.x>
- Gibson, J. J. (2015). *The ecological approach to visual perception*. Psychology Press.
- Gillespie, T. (2010). The politics of « platforms ». *New Media & Society*, 12(3), 347-364. <https://doi.org/10.1177/1461444809342738>
- Glaser, W. T., Westergren, T. B., Stearns, J. P. et Kraft, J. M. (2006). *Consumer item matching method and system* (United States brevet n° US7003515B1). <https://patents.google.com/patent/US7003515/en>
- Goldberg, D., Nichols, D., Oki, B. M. et Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61-70. <https://doi.org/10.1145/138859.138867>

- Gomez-Uribe, C. A. et Hunt, N. (2016). The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Transactions on Management Information Systems*, 6(4), 13:1-13:19. <https://doi.org/10.1145/2843948>
- Graus, D., Sappelli, M. et Manh Chu, D. (2018). Let Me Tell You Who You are – Explaining Recommender Systems by Opening Black Box User Profiles. Dans *FATREC workshop on Responsible Recommendation*. <https://piret.gitlab.io/fatrec2018/program/fatrec2018-graus.pdf>
- Gros, F. (2002). Sujet moral et soi éthique chez Foucault. *Archives de Philosophie*, 65(2), 229-237. <https://doi.org/10.3917/aphi.652.0229>
- Gros, F. (2005). Le gouvernement de soi. *Sciences Humaines, Hors série n°3*. https://www.scienceshumaines.com/foucault-derrida-deleuze-pensees-rebelles_fr_138.htm
- Grün, A. et Neufeld, X. (2021). Challenges Experienced in Public Service Media Recommendation Systems. Dans *Proceedings of the fifteenth ACM Conference on Recommender Systems*, 541-544. <https://doi.org/10.1145/3460231.3474618>
- Haider, J. (2016). The Shaping of Environmental Information in Social Media: Affordances and Technologies of Self-control. *Environmental Communication*, 10(4), 473-491. <https://doi.org/10.1080/17524032.2014.993416>
- Haim, M., Graefe, A. et Brosius, H.-B. (2018). Burst of the Filter Bubble? *Digital Journalism*, 6(3), 330-343. <https://doi.org/10.1080/21670811.2017.1338145>
- Hallinan, B. et Striphas, T. (2016). Recommended for you: The Netflix Prize and the production of algorithmic culture. *New Media & Society*, 18(1), 117-137. <https://doi.org/10.1177/1461444814538646>
- Harambam, J., Bountouridis, D., Makhortykh, M. et van Hoboken, J. (2019). Designing for the better by taking users into account: a qualitative evaluation of user control mechanisms in (news) recommender systems. Dans *Proceedings of the 13th ACM Conference on Recommender Systems*, 69-77. <https://doi.org/10.1145/3298689.3347014>
- Harambam, J., Helberger, N. et van Hoboken, J. (2018). Democratizing algorithmic news recommenders: how to materialize voice in a technologically saturated media ecosystem. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180088. <https://doi.org/10.1098/rsta.2018.0088>
- Hargreaves, E., Agosti, C., Menasché, D., Neglia, G., Reiffers-Masson, A. et Altman, E. (2018). Biases in the Facebook News Feed: a Case Study on the Italian Elections. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 806-812. <https://doi.org/10.1109/ASONAM.2018.8508659>
- Harper, F. M., Xu, F., Kaur, H., Condif, K., Chang, S. et Terveen, L. (2015). Putting Users in Control of their Recommendations. Dans *Proceedings of the 9th ACM Conference on Recommender Systems*, 3-10. <https://doi.org/10.1145/2792838.2800179>

- He, C., Parra, D. et Verbert, K. (2016). Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 56, 9-27. <https://doi.org/10.1016/j.eswa.2016.02.013>
- Heersmink, R. (2017). Distributed Cognition and Distributed Morality: Agency, Artifacts and Systems. *Science and Engineering Ethics*, 23(2), 431-448. <https://doi.org/10.1007/s11948-016-9802-1>
- Helberger, N. (2019). On the Democratic Role of News Recommenders. *Digital Journalism*, 7(8), 993-1012. <https://doi.org/10.1080/21670811.2019.1623700>
- Helberger, N. (2015). Public Service Media| Merely Facilitating or Actively Stimulating Diverse Media Choices? Public Service Media at the Crossroad. *International Journal of Communication*, 9(0), 17. <https://ijoc.org/index.php/ijoc/article/view/2875>
- Helberger, N., Karppinen, K. et D'Acunto, L. (2016). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, 21(2), 191-207. <https://doi.org/10.1080/1369118X.2016.1271900>
- Herlocker, J. L., Konstan, J. A., Borchers, A. et Riedl, J. (1999). An algorithmic framework for performing collaborative filtering. Dans *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 230-237. <https://doi.org/10.1145/312624.312682>
- Herlocker, J. L., Konstan, J. A. et Riedl, J. (2000). Explaining collaborative filtering recommendations. Dans *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, 241-250. <https://doi.org/10.1145/358916.358995>
- Hill, W., Stead, L., Rosenstein, M. et Furnas, G. (1995). Recommending and evaluating choices in a virtual community of use. Dans *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 194-201. <https://doi.org/10.1145/223904.223929>
- Hinchman, L. (1996). Autonomy, Individuality, and Self-Determination. Dans *Autonomy, Individuality, and Self-Determination* (p. 488-516). University of California Press. <https://doi.org/10.1525/9780520916890-034>
- Holtz, D., Carterette, B., Chandar, P., Nazari, Z., Cramer, H., & Aral, S. (2020). The Engagement-Diversity Connection : Evidence from a Field Experiment on Spotify. *Proceedings of the 21st ACM Conference on Economics and Computation*, 75-76. <https://doi.org/10.1145/3391403.3399532>
- Hutchins, E. L. (2001). Distributed Cognition. Dans N. J. Smelser et P. B. Baltes, *International Encyclopedia of the Social and Behavioral Sciences* (p. 2068-2072). Elsevier Press.
- Ihde, D. (1990). *Technology and the lifeworld: from garden to earth*. Indiana University Press.
- Jin, Y., Tintarev, N. et Verbert, K. (2018). Effects of personal characteristics on music recommender systems with different levels of controllability. Dans *Proceedings of the 12th ACM Conference on Recommender Systems*, 13-21. <https://doi.org/10.1145/3240323.3240358>

- Jugovac, M. et Jannach, D. (2017). Interacting with Recommenders: Overview and Research Directions. *ACM Transactions on Interactive Intelligent Systems*, 7(3), 10:1-10:46. <https://doi.org/10.1145/3001837>
- Kant, I. (2003). *Critique de la raison pratique* (J.-P. Fussler et M. Foessel, dir.). Éditions Flammarion. (Ouvrage original publié en 1788.)
- Kant, I. (2020). *Qu'est-ce que les Lumières ?* (J.-F. Poirier et F. Proust, trad.). Flammarion. (Ouvrage original publié en 1784.)
- Karakayali, N., Kostem, B. et Galip, I. (2018). Recommendation Systems as Technologies of the Self: Algorithmic Control and the Formation of Music Taste. *Theory, Culture & Society*, 35(2), 3-24. <https://doi.org/10.1177/0263276417722391>
- Katsaros, M., Yang, K. et Fratamico, L. (2022). Reconsidering Tweets: Intervening during Tweet Creation Decreases Offensive Content. *Proceedings of the International AAAI Conference on Web and Social Media*, 16, 477-487. <https://ojs.aaai.org/index.php/ICWSM/article/view/19308>
- Knijnenburg, B. P., Bostandjiev, S., O'Donovan, J. et Kobsa, A. (2012). Inspectability and control in social recommenders. Dans *Proceedings of the sixth ACM conference on Recommender systems*, 43-50. <https://doi.org/10.1145/2365952.2365966>
- Knijnenburg, B. P., Reijmer, N. J. M. et Willemsen, M. C. (2011). Each to his own: how different users call for different interaction methods in recommender systems. Dans *Proceedings of the fifth ACM conference on Recommender systems*, 141-148. <https://doi.org/10.1145/2043932.2043960>
- Knijnenburg, B. P., Sivakumar, S. et Wilkinson, D. (2016). Recommender Systems for Self-Actualization. Dans *Proceedings of the 10th ACM Conference on Recommender Systems*, 11-14. <https://doi.org/10.1145/2959100.2959189>
- Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H. et Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4), 441-504. <https://doi.org/10.1007/s11257-011-9118-4>
- Konstan, J. A. et Riedl, J. (2012). Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1), 101-123. <https://doi.org/10.1007/s11257-011-9112-x>
- Kotkov, D., Konstan, J. A., Zhao, Q. et Veijalainen, J. (2018). Investigating serendipity in recommender systems based on real user feedback. Dans *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, 1341-1350. <https://doi.org/10.1145/3167132.3167276>
- Krafft, T. D., Gamer, M. et Zweig, K. A. (2019). What did you see? A study to measure personalization in Google's search engine. *EPJ Data Science*, 8(1), 38. <https://doi.org/10.1140/epjds/s13688-019-0217-5>
- Lakoff, G. et Johnson, M. (2003). *Metaphors we live by*. University of Chicago Press.

- Latour, B. (1992). Where are the missing masses, sociology of a few mundane artefacts. Dans W. E. Bijker et J. Law (dir.), *Shaping technology/building society: studies in sociotechnical change*. MIT Press.
- Latour, B. (2000). Factures/fractures. De la notion de réseau à celle d'attachement. Dans A. MICOUD et M. PERONI (dir.), *Ce qui nous relie* (p. 189-208). Éditions de l'Aube. <http://spire.sciencespo.fr/hdl:/2441/516uh8ogmqildh09h61j7a8k5>
- Latour, B. (2006). *Petites leçons de sociologie des sciences* (Nouvelle éd.). la Découverte.
- Latzko-Toth, G., Bonneau, C. et Millette, M. (2016). Small Data, Thick Data: Thickening Strategies for Trace-based Social Media Research. Dans L. Sloan et A. Quan-Haase, *The SAGE Handbook of Social Media Research Methods* (p. 199-214). SAGE Publications Ltd. <https://doi.org/10.4135/9781473983847.n13>
- Le Béché, M. et Alloing, C. (2018). Au-delà des traces numériques visibles. Dans F. Liénard et S. Zlitni (dir.), *Réseaux sociaux, traces numériques et communication électronique*.
- Ledwich, M. et Zaitsev, A. (2019). Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization. *arXiv:1912.11211 [cs]*. <http://arxiv.org/abs/1912.11211>
- Leif, J. (1974). *Philosophie de l'éducation - Vocabulaire technique et critique de la pédagogie et des sciences de l'éducation* (vol. 4). Delagrave.
- Lejeune, C. (2014). *Manuel d'analyse qualitative : Analyser sans compter ni classer* (1re éd.). De Boeck.
- Light, B., Burgess, J. et Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, 20(3), 881-900. <https://doi.org/10.1177/1461444816675438>
- Lindberg, O. (2019, 12 avril). UX Evolutions: How Netflix Put UX at the Heart of its Ever-Changing Service. *Medium*. <https://medium.com/thinking-design/ux-evolutions-how-netflix-put-ux-at-the-heart-of-its-ever-changing-service-3c7e1b38c651>
- Mackenzie, C. (2014). Three dimensions of autonomy: A relational analysis. Dans A. Veltman (dir.), *Autonomy, oppression, and gender*. Oxford University Press.
- Mackenzie, C. (2019). Relational Autonomy: State of the art debate. Dans A. Armstrong, K. Green et A. Sangiacomo (dir.), *Spinoza and relational autonomy : being with others*. Edinburgh University Press.
- Mackenzie, C. et Stoljar, N. (dir.). (2000). *Relational autonomy: feminist perspectives on autonomy, agency, and the social self*. Oxford University Press.
- Marenko, B. et van Allen, P. (2016). Animistic design: how to reimagine digital interaction between the human and the nonhuman. *Digital Creativity*, 27(1), 52-70. <https://doi.org/10.1080/14626268.2016.1145127>
- Maslow, A. (1954). *Motivation and personality* (Harpers).

- May, T. (1994). The Concept of Autonomy. *American Philosophical Quarterly*, 31(2), 133-144. <https://www.jstor.org/stable/20014493>
- McLuhan, M. (2000). *Pour comprendre les médias : les prolongements technologiques de l'homme* (Nachdr.). Mame / Seuil.
- McNee, S. M., Riedl, J. et Konstan, J. A. (2006). Being accurate is not enough: how accuracy metrics have hurt recommender systems. Dans *CHI » 06 Extended Abstracts on Human Factors in Computing Systems*, 1097-1101. <https://doi.org/10.1145/1125451.1125659>
- Messing, S. et Westwood, S. J. (2014). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research*, 41(8), 1042-1063. <https://doi.org/10.1177/0093650212466406>
- Meyers, Diana T. (1989). *Self, society, and personal choice*. Columbia University Press.
- Meyers, Diana Tietjens. (2005). Decentralizing Autonomy: Five Faces of Selfhood. Dans J. Anderson et J. Christman (dir.), *Autonomy and the Challenges to Liberalism: New Essays* (p. 27-55). Cambridge University Press. <https://doi.org/10.1017/CBO9780511610325.004>
- Michaud, Y. (2000). Des modes de subjectivation aux techniques de soi : Foucault et les identités de notre temps. *Cités*, (2), 11-39. <https://www.jstor.org.proxy.bib.ucl.ac.be/stable/40620631>
- Milano, S., Taddeo, M. et Floridi, L. (2020). Recommender systems and their ethical challenges. *AI & SOCIETY*, 35(4), 957-967. <https://doi.org/10.1007/s00146-020-00950-y>
- Mill, J. S. (1956). *On Liberty*. Broadview Press.
- Minsky, M. (1986). *The society of mind*. Simon and Schuster.
- Moeller, J. et Helberger, N. (2018). *Beyond the filter bubble: Concepts, myths, evidence and issues for future debates*. <https://hdl.handle.net/11245.1/478edb9e-8296-4a84-9631-c7360d593610>
- Mori, M. (1970). Bukimi no tani (the uncanny valley). *Energy*, 7(4), 33-35.
- Negroponte, N. (1995). *Being digital* (1st ed). Knopf.
- Neher, A. (1991). Maslow's Theory of Motivation: A Critique. *Journal of Humanistic Psychology*. <https://doi.org/10.1177/0022167891313010>
- Newman, N., Fletcher, R., Schulz, A., Simge, A. et Nielsen, R. K. (2020). *Reuters Institute Digital News Report 2020* (p. 112).
- Nguyen, T. T., Hui, P.-M., Harper, F. M., Terveen, L. et Konstan, J. A. (2014). Exploring the filter bubble: the effect of using recommender systems on content diversity. Dans *Proceedings of the 23rd international conference on World wide web*, 677-686. <https://doi.org/10.1145/2566486.2568012>
- Norman, D. A. (1992). Design principles for cognitive artifacts. *Research in Engineering Design*, 4(1), 43-50. <https://doi.org/10.1007/BF02032391>

- Norman, D. A. (1993). Les artefacts cognitifs. Dans B. Conein, N. Dodier et L. Thévenot (dir.), *Les objets dans l'action : De la maison au laboratoire* (p. 15-34). Éditions de l'École des hautes études en sciences sociales. <http://books.openedition.org/editionsehess/9870>
- Norman, D. A. (2013). *The design of everyday things* (Revised and expanded edition). Basic Books.
- Oshana, M. (2016). *Personal Autonomy in Society*. Routledge. <https://doi.org/10.4324/9781315247076>
- Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you*. Penguin Books.
- Parra, D. et Brusilovsky, P. (2015). User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies*, 78, 43-67. <https://doi.org/10.1016/j.ijhcs.2015.01.007>
- Peraya, D. et Meunier, J.-P. (1999). Vers une sémiotique cognitive. In *Cognito*, 14, 1. <https://archive-ouverte.unige.ch/unige:17646>
- Postman, N. (2000). The humanism of media ecology. Dans *Proceedings of the media ecology association*, 1, 10-16.
- Proulx, S. (2015). La sociologie des usages, et après ? *Revue française des sciences de l'information et de la communication*, (6). <https://doi.org/10.4000/rfsic.1230>
- Pu, P., Chen, L. et Hu, R. (2011). A user-centric evaluation framework for recommender systems. Dans *Proceedings of the fifth ACM conference on Recommender systems*, 157-164. <https://doi.org/10.1145/2043932.2043962>
- Puschmann, C. (2019). Beyond the Bubble: Assessing the Diversity of Political Search Results. *Digital Journalism*, 7(6), 824-843. <https://doi.org/10.1080/21670811.2018.1539626>
- Rader, E. et Gray, R. (2015). Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. Dans *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 173-182. <https://doi.org/10.1145/2702123.2702174>
- Resnick, P., Garrett, R. K., Kriplean, T., Munson, S. A. et Stroud, N. J. (2013). Bursting your (filter) bubble: strategies for promoting diverse exposure. Dans *Proceedings of the 2013 conference on Computer supported cooperative work companion*, 95-100. <https://doi.org/10.1145/2441955.2441981>
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. et Riedl, J. (1994). GroupLens: an open architecture for collaborative filtering of netnews. Dans *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, 175-186. <https://doi.org/10.1145/192844.192905>
- Resnick, P. et Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56-58. <https://doi.org/10.1145/245108.245121>
- Revel, J. (2009). *Le vocabulaire de Foucault* ([Nouv. éd.]). Ellipses.

- Rich, E. (1979). User Modeling via Stereotypes. *Cognitive Science*, 3(4), 329-354. https://doi.org/10.1207/s15516709cog0304_3
- Rieder, B. (2020). *Engines of order: a mechanology of algorithmic techniques*.
- Rieder, B., Abdulla, R., Poell, T., Woltering, R. et Zack, L. (2015). Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring « We are all Khaled Said ». *Big Data & Society*, 2(2), 2053951715614980. <https://doi.org/10.1177/2053951715614980>
- Rogers, C. R. (1995). *On becoming a person: a therapist's view of psychotherapy*. Houghton Mifflin. (Ouvrage original publié en 1966.)
- Roth, C., Mazières, A. et Menezes, T. (2020). Tubes and bubbles topological confinement of YouTube recommendations. *PLOS ONE*, 15(4). <https://doi.org/10.1371/journal.pone.0231703>
- Rouvroy, A. et Berns, T. (2013). Gouvernementalité algorithmique et perspectives d'émancipation, Faced with algorithmic governmentality. *Réseaux*, (177), 163-196. <https://www.cairn.info/revue-reseaux-2013-1-page-163.htm>
- Salton, G., Wong, A. et Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620. <https://doi.org/10.1145/361219.361220>
- Schmitt, F. et Madelaine, N. (2019, 25 juin). *Laurent Frisch (Radio France): « La bataille de l'audio suivra celle des plates-formes vidéo »*. <https://www.lesechos.fr/tech-medias/medias/laurent-frisch-radio-france-la-bataille-de-laudio-suivra-celle-des-plates-formes-video-1032018>
- Schraube, E. (2009). Technology as Materialized Action and Its Ambivalences. *Theory & Psychology*, 19(2), 296-312. <https://doi.org/10.1177/0959354309103543>
- Seaver, N. (2019a). Captivating algorithms: Recommender systems as traps. *Journal of Material Culture*, 24(4), 421-436. <https://doi.org/10.1177/1359183518820366>
- Seaver, N. (2019b). Knowing Algorithms. Dans J. Vertesi et D. Ribes (dir.), *Digital STS* (p. 412-422). Princeton University Press. <http://nickseaver.net/papers/seaverMIT8.pdf>
- Shardanand, U. et Maes, P. (1995). Social information filtering: algorithms for automating word of mouth. Dans *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 210-217. <https://doi.org/10.1145/223904.223931>
- Shneiderman, B. (1989). Designin the user interface. Dans T. Forester (dir.), *Computers in the human context: information technology, productivity, and people* (1st MIT Press ed). MIT Press.
- Smets, A., Walravens, N. et Ballon, P. (2020). Designing Recommender Systems for the Common Good. Dans *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 276-278. <https://doi.org/10.1145/3386392.3399570>

- Sørensen, J. K. (2019). Public Service Media, Diversity and Algorithmic Recommendation: Tensions between Editorial Principles and Algorithms in European PSM Organizations. Dans *7th International Workshop on News Recommendation and Analytics*, 2554, 6-11.
- Star, S. L. et Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3), 387-420. <https://doi.org/10.1177/030631289019003001>
- Suchman, L. A. (2007). *Human-machine reconfigurations*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511808418>
- Sullivan, E., Bountouridis, D., Harambam, J., Najafian, S., Loecherbach, F., Makhortykh, M., Kelen, D., Wilkinson, D., Graus, D. et Tintarev, N. (2019). Reading News with a Purpose: Explaining User Profiles for Self-Actualization. Dans *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, 241-245. <https://doi.org/10.1145/3314183.3323456>
- Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton University Press. <http://www.jstor.org/stable/j.ctt7tbsw>
- Tallerås, K., Colbjørnsen, T., Oterholm, K. et Larsen, H. (2020). Cultural Policies, Social Missions, Algorithms and Discretion: What Should Public Service Institutions Recommend? Dans A. Sundqvist, G. Berget, J. Nolin et K. I. Skjerdingsstad (dir.), *Sustainable Digital Communities* (p. 588-595). Springer International Publishing. https://doi.org/10.1007/978-3-030-43687-2_49
- Tilleul, C. (2020). *Étude des relations entre les pratiques des jeunes adultes sur les réseaux socionumériques et le développement de leurs compétences en littératie médiatique*. Université Catholique de Louvain.
- Tintarev, N. et Masthoff, J. (2007). Effective explanations of recommendations: user-centered design. Dans *Proceedings of the 2007 ACM conference on Recommender systems*, 153-156. <https://doi.org/10.1145/1297231.1297259>
- Tricot, A., Plégat-Soutjis, F., Camps, J.-F., Amiel, A., Lutz, G. et Morcillo, A. (2003). Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH. Dans Desmoulins, C., Marquet, P., Bouhineau et D. (dir.), *Environnements informatiques pour l'Apprentissage humain 2003* (p. 391-402). ATIEF ; INRP. <https://edutice.archives-ouvertes.fr/edutice-00000154>
- Van Alstyne, M. et Brynjolfsson, E. (1996). Electronic communities: Global village or cyberbalkans? Dans J. DeGross, S. Jarvenpaa et A. Srinivasan (dir.), *Proceedings of the 17th International Conference on Information Systems* (p. 80-98).
- Varga, S. (2009). Life as Art: Concerning Some Paradoxes of an Ethical Concept. *Estetika: The European Journal of Aesthetics*, 46(1), 49-61. <https://doi.org/10.33134/eeja.51>
- Vargas, S. et Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. Dans *Proceedings of the fifth ACM conference on Recommender systems – RecSys '11*, 109. <https://doi.org/10.1145/2043932.2043955>

- Verbeek, P.-P. (2012). On Hubris and Hybrids: Ascesis and the Ethics of Technology. Dans P. Brey, A. Briggie et E. Spence (dir.), *The Good Life in a Technological Age* (1st Edition, p. 260-271). Routledge. <https://doi.org/10.4324/9780203124581-30>
- Verbeek, P.-P. (2016). Materializing Morality: Design Ethics and Technological Mediation. *Science, Technology, & Human Values*. <https://doi.org/10.1177/0162243905285847>
- Verbeek, P.-P. (2011). *Moralizing technology: understanding and designing the morality of things*. The University of Chicago Press.
- Vermersch, P. (2011). *L'entretien d'explicitation* (7e éd., nouv. éd. enrichie d'un glossaire). Esf éditeur.
- Vial, S. (2021). *Le design* (3e éd. mise à jour). Que sais-je ?
- Vrijenhoek, S., Kaya, M., Metoui, N., Möller, J., Odijk, D. et Helberger, N. (2021). Recommenders with a Mission: Assessing Diversity in News Recommendations. Dans *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (p. 173-183). Association for Computing Machinery. <https://doi.org/10.1145/3406522.3446019>
- Waytz, A., Heafner, J. et Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52, 113-117. <https://doi.org/10.1016/j.jesp.2014.01.005>
- Wiard, V., Briec, L. et Dufrasne, M. (2022). « The spy who loved me »: An qualitative exploratory analysis of the relationship between youth and algorithms. *Frontiers in Communication*. <https://doi.org/10.3389/fcomm.2022.778273>
- Wilkinson, D. (2018). Testing a recommender system for self-actualization. Dans *Proceedings of the 12th ACM Conference on Recommender Systems*, 543-547. <https://doi.org/10.1145/3240323.3240324>
- Wing, J. M. (2011). Computational thinking. Dans *2011 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 3-3. <https://doi.org/10.1109/VLHCC.2011.6070404>
- Winograd, T. et Flores, F. (1986). *Understanding computers and cognition: a new foundation for design*. Ablex Pub. Corp.
- Zuiderveen Borgesius, F., Trilling, D., Möller, J., Bodó, B., Vreese, C. H. de et Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*. <https://policyreview.info/articles/analysis/should-we-worry-about-filter-bubbles>