

## Fiche synthétique PLATFORM

**Objectif:** Notre action vise à examiner les impacts des algorithmes décisionnels sur les choix et préférences des consommateurs et leurs évolutions dans le temps.

**Données:** données Facebook et publicités de produits et d'information, tickets de caisse des clients TOTAL

### Questions scientifiques:

- Comparaisons de (i) comportements et de préférences sur (ii) l'acceptabilité de la publicité et sur (iii) l'évolution des habitudes de consommation en utilisant des méthodes directes (questionnaires) et indirectes (mouse tracking, adoption de promotions).
- Formalisation de modèles de préférences et de biais
- Design d'expériences contrôlées permettant de vérifier l'acceptabilité de la publicité en contrôlant le nombre de sujets, l'attrition, la compensation, et le biais (social, de statu quo, d'ancrage).
- Cadre juridique: RGPD/CNIL/Code de la Consommation

### De réfléchir à une façon synthétique de décrire votre Action/Atelier (nous souhaitons construire des fiches synthétiques de présentation des Actions MaDICS) :

#### - Enjeux de votre thématique

Notre action vise à créer et animer une communauté scientifique en informatique, marketing, économie comportementale et droit pour étudier le lien entre les résultats des algorithmes décisionnels et les déterminants comportementaux des consommateurs.

#### Enjeux

##### 1. Défis en économétrie et en économie comportementale

**Défi 1.1 : Développement d'un outil d'observation in situ des comportements des acteurs économiques et collecte de données massives de manière éthique permettant de comprendre les comportements économiques.** Il y a un large débat en économie comportementale sur les intérêts et les inconvénients des observations collectées en laboratoires comparées à celles collectées in situ. En effet, les expériences conduites en laboratoire permettent un contrôle maximal sur les conditions d'observation des comportements et de mieux isoler l'impact de la variable d'intérêt mais risque aussi de modifier le comportement des acteurs à cause de l'artificialité des conditions d'observation et rend difficile la généralisation des résultats (Levitt et List (2007a et b)). A l'inverse, l'observation des comportements *in situ* permet l'observation du comportement des acteurs en milieu écologique et donc une meilleure validité externe mais en contre parti cela augmente le bruit autour de l'expérimentation ce qui nuit à l'étude de l'impact de la variable d'intérêt (Camerer C.F. (2015), Falk et Heckman (2009)).

Le dispositif PLATFORM en permettant la collecte de données à grande échelle ainsi que le développement d'outil d'observation non intrusif de type AdImpact participe à trouver des solutions à cet arbitrage. En effet, la collecte massive de données permet de contrôler pour les conditions d'expérimentation. Par ailleurs, le développement de ces outils d'observation des comportements directement in situ et de communication avec les sujets ouvrent de formidables opportunités d'expérimentation.

**Défi 1.2 : Développement de techniques de comparaison des performances des outils d'économétrie et de machine Learning.** En économétrie, le dispositif PLATFORM permet de relever le défi du développement de techniques de comparaison des performances des outils d'économétrie et de machine Learning (par ex. la performance prédictive comparera les réseaux de neurones aux RUM, la

qualité d'estimation des consentements à payer opposera logit multinomiaux aux autres spécifications RUM estimées par les méthodes bayésiennes).

En économétrie, la littérature modélisant les choix en économie repose le plus souvent sur les **Discrete Choice Models** (DCM), fondés sur les modèles RUM (*Random Utility Models*) proposés par McFadden (McFadden (1974), McFadden (2001)). Les développements de ces modèles de régressions paramétriques (connus sous la nomination de modèles logit multinomiaux) ont intégré de multiples extensions statistiques / économétriques et comportementales. Par ailleurs les outils issus du Machine Learning permettent eux aussi d'analyser ces décisions de choix.

La multiplicité et la diversité des outils issus du Machine Learning questionnent la qualité des estimations et des prédictions. Quelques articles étudient les impacts de spécifications différentes sur une même situation de choix observés ou simulés (par ex. Munizaga and Alvarez-Daziano (2005), Fiebig et al. (2010), Bouscasse, Joly, and Peyhardi (2019)).

PLATFORM permettra d'étudier en termes de qualité des ajustements aux données, en termes de capacité prédictive, mais aussi en termes de qualité de production des indicateurs économiques et comportementaux dérivés des estimations, tels que les préférences des consommateurs, les consentements à payer ou les parts de marché. A terme, une grille d'analyse des différents modèles et outils de l'IA/ML pourra être proposée pour guider les choix des chercheurs en fonction de leurs objectifs et des situations de choix modélisées.

**Défi 1.3 : Utilisation des complémentarités entre les méthodologies d'IA et d'économétrie pour prédire et analyser les comportements d'achat des consommateurs.** Développement d'algorithmes de détection de changement permettant de suivre les comportements d'achats des consommateurs et d'étudier d'effectuer des appariements de consommateurs prenant en compte des caractéristiques inobservables et des relations de corrélations non étudiées ou trouvées par la théorie économique. L'économétrie quant à elle permettra d'analyser, une fois, les appariements effectués comment les variables économiques, tels que les attributs du produit, ou les réductions de prix, influencent le comportement des agents dans le temps.

## **2. Défis juridiques et éthiques**

### **Défi 2.1 : Les aspects juridiques et éthiques : regard amont**

En amont, dans le cadre de la mise en oeuvre du projet de recherche PLATFORM, la première question de recherche portera sur les modalités de collecte des données sensibles, ainsi que sur les modalités d'anonymisation des données collectées

([https://www.cnil.fr/sites/default/files/atoms/files/wp216\\_fr\\_0.pdf](https://www.cnil.fr/sites/default/files/atoms/files/wp216_fr_0.pdf)).

Une seconde question portera sur la mise en oeuvre complexe du régime applicable aux traitements de données à des fins de recherche :

([https://www.cnil.fr/sites/default/files/atoms/files/consultation\\_publicue\\_-\\_presentation\\_du\\_regime\\_juridique\\_applicable\\_aux\\_traitements\\_a\\_des\\_fins\\_de\\_recherche.pdf](https://www.cnil.fr/sites/default/files/atoms/files/consultation_publicue_-_presentation_du_regime_juridique_applicable_aux_traitements_a_des_fins_de_recherche.pdf))

### **Défi 2.2 : Les aspects juridiques et éthiques : regard aval**

Après analyse des données collectées par AdImpact sur la base des différentes enquêtes suivant l'envoi de différentes sortes de publicité, deux questions de recherche seront creusées :

La réglementation des publicités commerciales et des recommandations commerciales opérée dans le Code de la consommation modifié par la Loi Lemaire, dans plusieurs directives et règlements européens, dont le Règlement général sur la protection des données, est-elle adaptée aux différents types de données collectés et de traitements opérés à l'occasion de la mise en oeuvre de la publicité en ligne ? La collecte et le traitement des données sensibles doivent-ils être davantage encadrés, au regard des modalités de traitement de ce type de données, mises en lumière à l'occasion du projet de recherche PLATFORM ?

### **3. Défis en informatique**

**Défi 3.1 : Développement d'une méthodologie et d'outils permettant la reproductibilité des expériences contrôlées.** Les expériences contrôlées nécessitent le recrutement des participants, le déploiement des expériences et la génération de rapports. Nous développons AdImpact, une extension de AdAnalyst qui consiste en: un plugin permettant de générer des questionnaires pop-ups pour les observations directes, et une extension mouse tracking permettant de conduire des observations indirectes. AdImpact alimente une base de données contenant les résultats d'observations directes et indirectes de sujets. AdImpact fournit une librairie pour la reproductibilité des expériences, en particulier, la génération automatique de rapports d'expériences.

**Défi 3.2: Apprentissage de préférences et des valeurs.** L'observation du comportement décisionnel des individus permet de faire l'hypothèse que tel comportement est basé sur des préférences (et des valeurs) qui guident le comportement et que ces préférences puissent être "appris" par un système automatique. L'idée sous jacente est que une fois ces préférences apprises le système peut répliquer le comportement décisionnel de l'individu observé. Cependant il existe plusieurs problèmes. Le premier est l'hypothèse de comportement rationnel et en général guidée par des valeurs quand en réalité le comportement décisionnel peut être aléatoire, irrationnel ou guidé par des émotions. Il faut donc savoir limiter l'applicabilité de la démarche. Le deuxième est la nécessité d'utiliser de modèles qui seront forcément des approximations de la richesse des motivations d'un comportement décisionnel. Le troisième est l'hypothèse de répliquabilité parce que une décision satisfaisante (ou non) aujourd'hui peut ne pas l'être demain (en tout cas à une distance temporelle donnée). L'utilisation donc de cette démarche n'est pas généralisable, mais dépend du caractère du processus de décision qu'il faut appréhender et répliquer et du processus d'aide à la décision en cours. L'apprentissage des préférences et des valeurs pose également des problèmes techniques. Les préférences se présentent sous forme de "phrases" (preferential statements) qui sont modélisables sous forme de structures d'ordre et (si possible) de fonctions de valeur. Il faut donc des protocoles spécifiques ainsi que des méthodes d'apprentissage spécifiques pour cette typologie de données. Il faut savoir distinguer entre l'apprentissage de modèles et l'apprentissage de paramètres (étant donné une hypothèse de modèle). Enfin il faut faire attention sur le fait que les préférences étant subjectives, il n'existe pas de "benchmarks" pour valider une procédure d'apprentissage. Au contraire l'apprentissage de préférences peut avoir comme conséquence la révision de valeurs et préférences existantes.

#### **- Solutions existantes (ce que l'on sait (bien) faire)**

La réalisation de notre action nécessite l'établissement d'un dialogue entre experts en sciences et technologies de l'information et en sciences de la consommation. D'une part, les informaticiens devront développer des approches permettant de révéler les préférences des individus. D'autre part, les comportementalistes devront concevoir des expériences contrôlées selon des principes établis en psychologie cognitive et en économie comportementale. Les solutions existantes savent bien faire l'un et l'autre mais pas les deux. De plus, l'élicitation de préférences se fait soit avec des observations directes

(questionnaires) ou indirectes (mouse tracking). Elles ne savent bien faire l'un et l'autre mais pas les deux. Enfin, les approches existantes ne confrontent pas les résultats d'expériences contrôlées à des résultats de fouille de données historiques. Elles savent faire les deux mais séparément.

**- Questions ouvertes : verrous, thèmes scientifiques existants et nouveaux, thèmes scientifiques structurants et/ou émergents, prospectives scientifiques Qu'est ce qui pourrait être produit par votre Action/Atelier dans les 2 prochaines années (état des lieux sur une de ces questions ?, panorama de solutions ?, jeux de données, mise en réseau de partenaires pour montage de projets,...) ?**

Notre thème scientifique structurant est: développer une méthodologie de travail entre chercheurs en marketing/l'économie comportementale, et chercheurs en techniques algorithmiques pour la collecte et l'analyse de grands volumes de données sur les préférences et comportements des utilisateurs/les approches de recommandation.

Nous produisons **un lexique commun** pour communiquer avec le même vocabulaire.

Nous développons **AdImpact**, un plug-in Facebook utilisable par tous, qui enregistre et agrège les informations sur les publicités de produits et d'information. AdImpact permettra aussi de réaliser des expériences contrôlées avec des observations directes (questionnaires) et indirectes (mouse tracking).

Nous développons également **un guide juridique**.

**Recueillir vos éléments de réponse, avis et réflexions que nous ferons remonter au CNRS suite aux 2 questions suivantes que nous pose le CNRS et qui sont relatives à la crise liée au Covid-19 :**

- en quoi cette crise change nos objectifs scientifiques et/ou nos objets d'études ? Cette crise met-elle en évidence de nouvelles problématiques scientifiques ? Va-t-elle renforcer l'interdisciplinarité ?

- en quoi cette crise change nos façons de travailler, notre mode de fonctionnement et notre organisation au travail ?

**Nous avons déposé un projet EMERGENCE pour recueillir des informations concernant COVID-19 à travers des campagnes d'information lancées sur AdImpact.**

### **Bibliographie.**

Fürnkranz J., Hüllermeier E. (eds) Preference Learning. Springer, Berlin, Heidelberg, 2010.

<http://www.mdai.cat/mdai2015/mdai2015.Plenary.Hullermeier.pdf>

<https://da2pl.cs.put.poznan.pl/>

Berry, S. T., & Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, 82(5), 1749-1797.

Bouscasse H, Joly I, Peyhardi J (2019). "A new family of, qualitative choice models: An application of reference models to, travel mode choice." *Transportation Research Part B*, 74-91.

Camerer C.F. (2015). The Promise and Success of Lab Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List. In *Handbook of Experimental Economic Methodology*, Oxford University Press.

Deaton A. (2016). Measuring and Understanding Behavior, Welfare, and Poverty. *American Economic Review*, 106(6), 1221-1243.

Fiebig D, Keane M, Louviere J, Wasi N (2010). "The generalized, multinomial logit model: accounting for scale and coefficient, heterogeneity." *Marketing Science*, 29:3, 393-421.

Levitt S.D, List J.A. (2007a). Viewpoint: On the Generalizability of Lab Behaviour to the field. *Canadian Journal of Economics*, vol. 40, n° 2, pp. 347-370.

Levitt S.D, List J.A. (2007b). What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World? *Journal of Economic Perspectives*, vol. 21, n° 2, pp. 153-174.

McFadden D (1974). "The measurement of urban travel demand.", *Journal of Public Economics*, 3 (4), 303-328.

McFadden D (2001). "Economic Choices." *The American Economic Review*, 91 (3), 351-378. ISSN 00028282, <URL: <http://www.jstor.org/stable/2677869>>.

Munizaga M, Alvarez-Daziano R (2005). "Testing mixed logit and, probit models by simulation." *Transportation Research Record*: *Journal of the Transportation Research Board*, 1921, 53-62.

Omar Al-Ubaydli & John A. List & Dana L. Suskind, 2017. "What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results," *American Economic Review*, *American Economic Association*, vol. 107(5), pages 282-286, May.

Drechsler, Jörg. 2011. *Synthetic Datasets for Statistical Disclosure Control: Theory and Implementation*. Vol. 201. Springer Science & Business Media. <https://www.springer.com/gp/book/9781461403258#reviews>.

Charest, Anne-Sophie. 2011. "How Can We Analyze Differentially-Private Synthetic Datasets?" *Journal of Privacy and Confidentiality* 2 (2). <https://doi.org/10.29012/jpc.v2i2.589>.