Ecole doctorale LECLA
Projet de thèse « Contrat doctoral UB/UFC »
ANNEE 2022

Porteur du projet : BRACHOTTE Gilles, gilles.brachotte@u-bourgogne.fr
Section CNU : 71(Sciences de l'Information et de la Communication)

Titre du projet de thèse : La manipulation de l'information sur Twitter : détection, circulation et stratégies

1. Contexte scientifique et état de l'art

Contexte : Le réseau social Twitter, véritable chambre de résonnance qui favorise l'immédiateté et la rapidité de propagation de l'information, sert de relais pour les autres médias (Brachotte & Frame, 2015 ; 2016) dans une logique polymédia (Madianou & Miller, 20123) offrant à tout un chacun une tribune médiatique de libre expression de son opinion sans censure ni contrôle apparent. Les discours et les informations se retrouvent amplifiés et polarisés par l'effet « bulle » de Twitter. Celui-ci facilite, entre autres, la visibilité et la propagation de l'information manipulée à l'image des fausses nouvelles. Cette potentialité de Twitter est exploitée, à l'insu des twittos, par des personnes et des communautés qui font appel à la technique informatique comme les robots sociaux à des fins d'influence en effectuant des opérations automatisées tant dans leur modalité (répétitives, prédéfinies, etc.) que dans leur intentionnalité (polarisantes, trompeuses, manipulatoires). Plusieurs études ont montré l'utilisation de robots sociaux à des fins politiques en mettant en œuvre des campagnes de désinformation. Cela a été le cas lors des élections présidentielles françaises en 2017 contre le candidat à la présidence de la république Emmanuel Macron (Ferrara, 2017 ; Frame et al., 2021), de la campagne électorale américaine (Bessi et Ferrara, 2016 ; Kollanyi, 2016), de l'élection fédérale allemande en 2017 (Brachten et al., 2017 ; Bender et Oppong, 2017) ou encore dans la politique vénézuélienne en général (Forelle et al., 2015), la guerre civile syrienne (Abokhdair et al., 2015) et depuis 2015 dans le conflit entre l'Ukraine et la Russie (Hegelich et Janetsko, 2016 ; Smart et al, 2022).

Positionnement du projet par rapport à l'état de l'art : Depuis les années 2010, les robots (au sens du programme informatique capable de publier automatiquement du contenu) présentent un intérêt grandissant du fait de leur rôle potentiel dans la diffusion et/ou la manipulation de l'information. Plusieurs botnets de plusieurs centaines de milliers de comptes ont ainsi pu être identifiés par le passé (Echeverria et al., 2017 ; Besel et al., 2018). On peut distinguer différentes catégories de robots : a) les robots au comportement bénin comme ceux facilitant la diffusion d'informations telles que les actualités (Lokot et al., 2016), les événements climatiques et environnementaux (Haustein et al., 2016) ou encore les *chatbots* (Rodrigo et al., 2012) capables d'interagir avec l'utilisateur, b) les robots au comportement malicieux qui cherchent à profiter des dispositifs sociotechniques comme les réseaux sociaux numériques (RSN) afin de servir une personne ou un groupe de personnes sans clairement indiquer leurs intentions. Plusieurs stratégies peuvent alors être mise en œuvre pour parvenir à leur fin. La plus basique consiste à réaliser des attaques de spam (Wang et al. 2010). D'autres sont plus évoluées et requièrent le déploiement de robots sociaux, dont les profils et comportements sont proches des utilisateurs humains (Cresci et al. 2020) facilitant ainsi l'interaction sans pourtant être identifiés comme des programmes informatiques automatisés. Parmi ces stratégies, on retrouve *l'astroturfing* (Ratkiewicz et al. 2011), la *misdirection*, ou l'écran de fumée (Abokhodair et al. 2015). *L'astroturfing* consiste à simuler un mouvement d'approbation (ou de désapprobation) afin de donner l'impression qu'une majorité adhère à l'idée diffusée, dans le but d'influencer les utilisateurs du réseau social. La *misdirection* et l'écran de fumée sont deux concepts proches, qui consistent à se servir de sujets populaires tout en discutant d'autre chose, dans le but de distraire les utilisateurs du point principal du sujet d'origine. L'écran de fumée va créer une discussion ayant un lien avec le contexte, tandis que la *misdirection* discutera d'un tout autre sujet.

2. Argumentaire technique et scientifique

Description des hypothèses de recherche et des objectifs : Les techniques de détection de robots actuelles consistent principalement à recourir à des techniques de machine learning nécessitant une phase d'apprentissage sur des données annotées et/ou à se concentrer sur une vérification individuelle de comptes. Or, le mimétisme des robots avec l'humain impose de faire évoluer les techniques de détection qui doivent être plus évolutive et considérer le réseau dans son ensemble afin d'identifier les comportements similaires et révéler alors des armées de robots dont la stratégie est potentiellement globale et supervisée. Nous posons comme hypothèses de recherche que sur Twitter : a) existe des armées de robots présentant des régularités ou des logiques de comportements similaires (dans la production des tweets et dans leur structure intrinsèque) puisqu'ils sont pilotés par un algorithme (même s'il y a une partie aléatoire) (Kollanyi 2016), b) les robots au sein d'une armée peuvent posséder des fonctions spécifiques, complémentaires et hiérarchisées au service d'une action coordonnée et d'un objectif commun, comme par exemple des robots « éclaireurs » qui détectent des thématiques particulières ; robots « fantassins » qui servent massivement à détourner ou à faire émerger une information ; robots « infiltrés » visibles autour de thématiques ciblées qui servent de relais légitimant, c) les robots sociaux utilisés à des fins politiques sont l'une des formes les plus importantes de robots sociaux de par leur capacité à publier du contenu idéologique ou de participer aux débats politiques affectant toute l'écologie médiatique, et enfin, corollaire de l'hypothèse précédente, d) la conception et la programmation des robots sociaux sont principalement l'apanage d'activistes, de gouvernements, de groupes d'individus engagés et/ou militants souhaitant diffuser et promouvoir des campagnes, des idées et des positions politiques (Smart et al 2022, Abdine 2022).

D'un point de vue méthodologique, le doctorant mettre en œuvre une mix-méthode (quantitatif et qualitatif) et profitera de la plateforme à haute performance du projet COCKTAIL qui permet le stockage et la collecte de tweets à partir de critères définis (mots clefs, @, #). Cette plateforme fournit également un outillage algorithmique permettant, entre autres, de déterminer des communautés, potentiellement polarisées, de comptes et/ou de hashtags et/ou de mots clefs. De plus, le projet COCKTAIL fournira deux corpus déjà constitués de plusieurs millions de tweets à propos de la vaccination contre la COVID-19 et les

élections présidentielles française 2017. Ces deux corpus présentent un terrain favorable à la présence de robots sociaux et à la manipulation de l'information. Dans un premier temps, le doctorant identifiera les robots sociaux à l'aide d'outils existants comme Botometer mais également à partir d'une décomposition tensorielle développée dans COCKTAIL. Dans un deuxième temps, via une analyse qualitative, il confirmera l'existence des robots et pourra analyser leur influence sur la circulation de l'information. Enfin, il proposera des méthodes et des stratégies pour contrer ces manipulations d'informations. L'ensemble pourra être testé sur un jeu de données actualisé qui reste à définir en fonction de l'actualité.

3. Objectifs et résultats escomptés

L'objectif principal de la thèse est de comprendre le rôle et l'usage des robots sociaux et des armées de robots (botnet) dans la diffusion et la manipulation de l'information sur Twitter. L'étude de l'état de l'art a montré que peu de travaux décryptent les stratégies de fonctionnement des armées de robots. Les travaux se concentrent majoritairement sur la détection des comptes pris individuellement et analyse le rôle et la stratégie unique du robot.
Cette thèse en Sciences de l'Information et de la Communication a pour ambition de mieux :

- comprendre les phénomènes sociaux, politiques, économiques, etc. pouvant être affectés par l'utilisation d'agents conversationnels afin d'essayer de déclencher des effets de masse, notamment dans un contexte de « désinformation » (contexte électoral, alertes alimentaires/sanitaires, dénonciations…) ;
- comprendre et caractériser l'utilisation et le fonctionnement d'agents conversationnels dans des campagnes de « désinformation » et les stratégies déployées pour éviter leur détection ;
- comprendre et caractériser les dynamiques socio-techniques de la circulation d'informations sur Twitter.

Les résultats attendus permettront de comprendre les logiques fonctionnelles des armées de robots et leur(s) rôle(s) dans les stratégies d'influence et de manipulation de l'information afin de mettre en place des stratégies de réponse suite à des campagnes de désinformation sur Twitter.
De plus, les résultats permettront de déterminer comment les armées de robots s'intègrent dans le réseau social et leur positionnement dans la structuration communautaire thématisée de Twitter. La thèse vise également à analyser si l'activité des robots a une influence sur la topologie du réseau social et si ceux-ci participent à une reconfiguration du réseau social.

4. Laboratoire de rattachement et Insertion du projet dans les axes de recherche du labo

Laboratoire de rattachement CIMEOS EA 4177, axe « Transition socio-écologique, territoires et espaces publics », Université de Bourgogne, Campus de Dijon (http://cimeos.u-bourgogne.fr)

5. Partenariats éventuels, environnement scientifique

Au sein de la MSH de Dijon et sous l'autorité du responsable scientifique du programme de recherche « COCKTAIL », le doctorant(e) fera partie d'une équipe pluridisciplinaire.

6. Bibliographie indicative (courte)

Abokhodair, N., Yoo, D., & McDonald, D. W. (2015). Dissecting a social botnet: Growth, content and influence in Twitter. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, 839-851.
Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2012). Detecting automation of twitter accounts: Are you a human, bot, or cyborg?. *IEEE Transactions on dependable and secure computing*, *9*(6), 811-824.
Cresci, S. (2020). A decade of social bot detection. *Communications of the ACM*, *63*(10), 72-83.
Ferrara E. (2017). Disinformation and social bot operations in the run up to the 2017 french presidential election. https://doi.org/10.5210/fm.v22i8.8005)
Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A., 2016. The rise of social bots. *Communications of the ACM 59*, 96–104.
Frame, A., & Brachotte, G. (2021). Viral Tweets, Fake News and Social Bots: Post-Truth PR and the French General Elections 2017. *Epistémè*, 25, 53–78.
Gillet, A., Leclercq, É., & Cullot, N. (2022). Multi-level optimization of the canonical polyadic tensor decomposition at large-scale: Application to the stratification of social networks through deflation. *Information Systems*, 102142.
Lokot, T., & Diakopoulos, N. (2016). News Bots: Automating news and information dissemination on Twitter. *Digital Journalism*, *4*(6), 682-699.
Stieglitz, S., Brachten, F., Ross, B., & Jung, A. K. (2017). Do social bots dream of electric sheep? A categorisation of social media bot accounts. *arXiv preprint arXiv:1710.04044*.
Smart, B., Watt, J., Benedetti, S., Mitchell, L., Roughan, M. #IStandWithPutin versus #IStandWithUkraine : The interaction of bots and humains in discussion of Russia / Ikraine war, https://arxiv.org/pdf/2208.07038.pdf consulté le 31/10/200
Wang, A. H. (2010). Detecting spam bots in online social networking sites: a machine learning approach. In *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, Berlin, Heidelberg, 335-342.
Woolley, S. C. & Howard, PN. (2016). Political Communication, Computational Propaganda, and Autonomous Agents. *International Journal of Communication,* 10(2016), 4882–4890

7. Calendrier prévisionnel

La durée du doctorat est de 36 mois à temps plein à compter du 1er octobre 2023.

Première année : état de l'art et positionnement du projet

Deuxième année : confirmation des hypothèses et des objectifs, développement de la méthodologie, traitement des données et publications

Troisième année : travail de terrain, traitement des données, publications et soutenance.