Décembre 2016

# Création d'une plateforme scientifique pour le développement de la transparence et de la responsabilité des algorithmes et des données « *TransAlgo* »

## Résumé


Suite à une saisine d'Axelle Lemaire, Secrétaire d'Etat chargée du numérique et de l'Innovation, le Conseil Général de l'Economie[1] (CGE) a rédigé le rapport « Modalités de régulation des algorithmes de traitement des contenus ». L'une des recommandations formulées vise à la mise en place d'une plateforme scientifique collaborative destinée à favoriser, d'une part le développement d'outils logiciels et de méthodes de tests d'algorithmes « responsables et transparents », et d'autre part la promotion de leur utilisation.

**Inria se propose de porter une telle plateforme, dénommée *TransAlgo*, qui contribuera à développer une culture et un savoir-faire pour une production, une analyse algorithmique et une valorisation des données *responsables et éthiques*. *TransAlgo* aidera aussi à diffuser les bonnes pratiques auprès des services de l'Etat, des industriels et des citoyens**.

Cette plateforme associera dès le début d'autres acteurs académiques, en particulier l'Institut Mines-Telecom et le CNRS, dans des modalités en cours de finalisation. Elle sera développée en coopération avec le Conseil National du Numérique (CNNum), la Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes (DGCCRF) et la Direction Générale des Entreprises (DGE), en veillant au respect des missions et rôles de chacun.

**Cette plateforme sera une première en Europe**. Avec l'émergence et le développement des technologies Big Data et de l'Intelligence Artificielle, ces questions deviennent capitales pour le citoyen, pour les pouvoirs publics et pour le monde de l'innovation et de la recherche. Outre-atlantique les réflexions sont d'ores et déjà bien lancées ; pour exemple la publication du plan stratégique en recherche et développement en Intelligence Artificielle de la Maison Blanche[2] (octobre 2016) ainsi que l'initiative "Explainable AI"[3] lancée par la DARPA (Defense Advanced Research Projects Agency) en août 2016 pour mieux comprendre le comportement des algorithmes et les aligner avec les règles éthiques et légales. Un récent rapport de la Maison Blanche (mai 2016)[4] ainsi que celui de la Federal Trade Commission[5] (FTC – Janvier 2016) attirent par ailleurs l'attention sur les risques d'exclusion et d'impact sur les droits civiques qui pourraient être occasionnés par les technologies Big Data si nous ne faisons pas attention à la maitrise de leurs conditions d'utilisation.

**Le lancement de la plateforme TransAlgo contribue à ce que puisse s'appliquer l'exigence de transparence et de responsabilité de acteurs de l'économie numérique, introduit par la loi pour une République numérique.**


[1] http://www.economie.gouv.fr/cge/actualites
[2] https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf
[3] http://www.darpa.mil/program/explainable-artificial-intelligence
[4] https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf
[5] https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf


Domaine de Voluceau - Roquencourt B.P. 105
78153 Le Chesnay Cedex France
Tél. : +33 (0)1 39 63 55 11
Fax : +33 (0)1 39 63 55 30

www.inria.fr

## Objectifs et services rendus

**Les méthodes et les outils techniques liés à la responsabilité et à la transparence des algorithmes sont un sujet complexe et multiforme**. Les propriétés que l'on souhaite vérifier, par exemple l'équité, la non-discrimination ou la loyauté... incluent une part importante de **subjectivité et de choix de conception** dépendant des cas d'usage et des contextes qui rend leur **spécification complexe et difficile**. Les challenges scientifiques sont nombreux et très peu de travaux de recherche sur le sujet sont encore disponibles[6].

C'est dans ce cadre général que se situent les trois principaux objectifs de *TransAlgo*.

**Le premier est d'encourager la conception d'algorithmes de traitement de données « responsables et transparents par construction** (on parle de « *responsible-by-design*») ». Un algorithme est dit « responsable » s'il respecte les lois (e.g. la confidentialité de certaines données, la non-discrimination par ses critères), et s'il se conforme à certaines règles éthiques (e.g. la neutralité). En outre, un algorithme transparent se doit de faciliter la vérification de sa responsabilité, par exemple, en ouvrant son code, en explicitant la provenance des données qu'il a utilisées, et celles qu'il produit, en expliquant ses résultats, ou en publiant des traces de ses calculs.

**Le second objectif de *TransAlgo* est d'aider à la vérification et au test de ces algorithmes**, notamment à vérifier que ces derniers se comportent comme ils sont tenus de le faire (légalement) et comme ils déclarent le faire (« loyalement »). Il convient de distinguer deux cas, les algorithmes dont le code est ouvert aux autorités et les algorithmes dont le code ne l'est pas (boîte noire). Dans ce second cas, un défi supplémentaire s'ajoute dans la mesure où les autorités doivent disposer (via *TransAlgo*) des méthodes, des outils et des jeux de données adéquats pour « entrouvrir la boîte noire ». Cela consiste par exemple à stimuler l'algorithme par des données ou des profils en entrée et à observer les réponses en sortie. Une finalité est de permettre aux autorités de régulation de confondre les auteurs d'algorithmes « irresponsables » ne respectant pas la loi, de manière intentionnelle ou non.

**Le troisième objectif de *TransAlgo* est d'aider à la diffusion de savoir-faire et de bonnes pratiques** auprès des services de l'Etat, des industriels et des citoyens. Elle doit aussi permettre de traiter les attentes de ces différents acteurs, attentes qui nourriront les sujets de recherche à étudier.

Du fait de la dualité des données et des algorithmes (qui en assurent l'analyse et la gestion), il est essentiel de considérer les outils et les algorithmes de transparence et de responsabilité adressant à la fois les sujets relatifs à la qualité, la typologie (sensibles ou personnelles), la provenance, la représentativité des données ; mais aussi les questions qui se posent pour la traçabilité, le contrôle, le caractère explicable, l'usage (paramétrage et critères) et la réutilisabilité des algorithmes. Dans le cas des données personnelles, il est essentiel de développer les méthodes de protection de la vie privée par des approches de "Privacy-by-design" incluant des techniques comme "Differential privacy"[7].

Le développement des méthodes responsables et éthiques pour la gestion et l'analyse des données mêle diverses **compétences pluri-disciplinaires** comme les statistiques et l'apprentissage automatique, les télécommunications, les bases de données, la visualisation des

[6] Nous citons, par exemple, deux workshops récents qui représentent une communauté scientifique naissante en cours de cristallisation autour de ces sujets:

- "Fairness, Accountability and Transparency in Machine Learning" 2016, http://www.fatml.org/schedule/2016
- "Data and Algorithmic Transparency" 2016, http://datworkshop.org/#tab_home

[7] Méthodes puissantes de protection des données personnelles qui introduisent du bruit statistique et des aléas rendant difficile l'identification des individus.

données, la cryptographie et la protection des données, l'économie des services numériques, la régulation, la sociologie computationnelle, etc.

Il est essentiel de traiter différents cas d'usage en explicitant les critères de conformité à la réglementation afin de spécifier les critères de mesure de la transparence incluant la robustesse au détournement, la mesure du biais inhérent ou encore la traçabilité du raisonnement automatique pour être en mesure d'identifier les responsabilités dans les situations décisionnelles à fort impact.

*TransAlgo* ne se limitera pas à la mesure de la transparence des plateformes du web mais considérera la transparence des algorithmes de façon générale, quels que soient leurs supports d'exécution, par exemple les Smartphones. Un exemple emblématique de cette catégorie est l'étude *Mobilitics*[8], une collaboration de recherche entre la CNIL et Inria pour mieux comprendre les mécanismes d'accès aux données personnelles par les applications des Smart Phones. Cette étude a révélé la non loyauté de certains algorithmes en libre circulation, permettant de questionner la responsabilité de leurs propriétaires.

**En résumé**, la plateforme *TransAlgo* sera à la fois :

1. **Centre de ressources** : liens vers des projets pertinents, des outils, des travaux, des expériences, des points de vue, des initiatives internationales, etc. Le but est ici de centraliser les efforts de la communauté scientifique sur le sujet, d'entretenir les liens et les échanges avec d'autres initiatives similaires comme le « *Data Transparency Lab*[9] » (DTL). La plateforme apportera des ressources comme des algorithmes et des données et un espace d'expérimentation logicielle pour la mesure des différents aspects de la transparence ;
2. **Instrument d'incitation pour le développement de nouveaux outils et méthodes** via des appels à projets de recherche ciblés, des challenges, des expérimentations, etc ;
3. **Moyen de promotion de ces outils et méthodes auprès des pouvoirs publics**, des industriels et des citoyens et d'aide à la transformation des systèmes algorithmiques existants.

Il convient de souligner que, pour garantir son indépendance et la liberté des recherches dont elle se fera le porteur ou le relais, **la plateforme scientifique *TransAlgo* ne sera en aucun cas en charge du contrôle réglementaire des algorithmes ou de l'utilisation des données**. Elle proposera par contre une offre d'études, d'outils et de services à l'ensemble des acteurs concernés.

[8] https://team.inria.fr/privatics/mobilitics/
[9] http://www.datatransparencylab.org

## Acteurs et gouvernance

**Porteur(s) et partenaires académiques**:

Inria opèrera la plateforme et jouera un rôle de catalyseur de la dynamique scientifique avec d'autres partenaires académiques, notamment le CNRS et l'Institut Mines-Telecom. Outre l'expertise scientifique, Inria apportera l'aide au développement logiciel.

**Autres partenaires :**

Un enjeu majeur de *TransAlgo* est de répondre à la préoccupation légitime des utilisateurs de comprendre les algorithmes qui gouvernent une partie importante de leurs activités, ou à défaut de pouvoir s'appuyer sur un tiers de confiance. Dans cet esprit, *TransAlgo* sera développée en coopération étroite avec le Conseil National du Numérique (CNNum) et la Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes (DGCCRF).

- Le CNNum s'est vu confier une mission de recensement et d'objectivation de la situation actuelle sur les pratiques des plateformes et de leurs paramètres de différentiation comme par exemple la réputation. Il mettra en place à cet effet un dispositif contributif d'expérimentation de l'évaluation de ces pratiques. Il mène plus globalement une action d'enrichissement de la réflexion générale via l'organisation régulière de débats citoyens, et d'aide à une acceptabilité consentie et éclairée notamment par des expressions de besoin.
- Au-delà du CNNum et la DGCCRF pourront solliciter la communauté scientifique par leurs attentes et leurs besoins. Ainsi, les résultats scientifiques et techniques suffisamment mûrs issus de *TransAlgo* pourront alimenter les outils techniques de contrôle et de régulation, objet d'une autre recommandation du rapport du CGE.

**Autres acteurs potentiels**:

- Académiques : universités et écoles, organismes de recherches (notamment dans le cadre de l'alliance Allistene[10]) qui contribuent au développement des recherches, des enseignements et des compétences
- Industriels : grandes entreprises, PME et start-ups qui apportent des cas d'usages, des données, des problématiques mais aussi qui viennent tester leurs composants logiciels dans le cadre d'une démarche volontariste
- Autorités de régulation et instances officielles: ANSSI[11], CERNA[12],... qui apportent leurs compétences, leurs points de vue et leurs demandes dans la conception technique, scientifique et éthique. Il sera en particulier proposé à la CNIL[13] d'être associée étroitement aux travaux autour de *TransAlgo*.
- Utilisateurs et consommateurs : FING[14], associations de consommateurs comme "UFC-Que Choisir"[15], "60 Millions de Consommateurs"[16]... qui contribuent à la réflexion générale, enrichissent le débat citoyen et œuvrent à l'acceptabilité, en particulier par une expression de besoins.

[10] https://www.allistene.fr/
[11] http://www.ssi.gouv.fr/
[12] https://www.allistene.fr/cerna-2/
13 https://www.cnil.fr
[14] http://www.fing.org
[15] https://www.quechoisir.org/
[16] http://www.60millions-mag.com/

## Calendrier et actions

Le calendrier ci-dessous est donné à titre indicatif, il sera affiné au fur et à mesure du développement de la plateforme, des avis et décisions du comité de pilotage et des fonds disponibles.

| I - Phase de démarrage couvrant les 4 premiers mois |
|---|
| **1- Déploiement de la plateforme technique**<br>• Achat de matériel<br>• Hébergement technique (souverain) de la plateforme *TransAlgo*<br>• Recrutement d'un ingénieur qui va déployer la plateforme et en assurer le maintien en condition opérationnelle<br>• Recrutement d'un ingénieur de développement logiciel pour l'industrialisation des résultats de la recherche pour leur mise à disposition sur la plateforme |
| **2- Alimentation en contenu (l'accès à ce contenu se fait à travers un portail web)**<br>• Contenu logiciel : collecter les premières briques logicielles académiques existantes et les mettre à disposition[17] avec la description des conditions d'utilisation pour les services de l'état, les industriels, les associations de consommateurs,...<br>• Contenu documentaire : collecter des rapports, des livres blancs, des articles scientifiques, des conférences, des initiatives internationales… |
| **3- Animation de la communauté scientifique**<br>• Lancer un premier appel à contributions voire un appel à projets de recherche. |

[17] La mise à disposition peut prendre deux formes: soit par téléchargement soit via un accès par web-services. Pour la phase de démarrage, seul l'accès par téléchargement sera possible et la plateforme se comportera comme un centre de ressources.

**II - Phase de montée en puissance à partir du 5ème mois**

(Cette montée en puissance sera reconsidérée sur une base annuelle en terme de ressources techniques et humaines en fonction de l'activité scientifique, des expressions de besoin et des cas d'usage fourni par le comité de pilotage)

**1- Extension technique la plateforme**

- Ajout de serveurs (de calcul et de stockage)

**2- Enrichissement du contenu de la plateforme**

- Intégration et mise à disposition de nouvelles briques logicielles éventuellement disponibles via des web-services, ces briques peuvent être les outils de mesure de la transparence comme des outils de génération automatique de profils ou autres données utiles pour l'exécution des tests. Implémentation des principales fonctionnalités, services et fonctions-clés spécifiés et priorisés par le Comité de pilotage, sur proposition du Comité d'experts
- Contenu documentaire: création de MOOCs pour des publics variés (pouvoirs publics, industriels, citoyens) dans le domaine de la responsabilité et de la transparence des données et des algorithmes
- Recrutement d'ingénieurs de développement logiciel

**3- Animation de la communauté scientifique**

- Lancement d'un nouvel appel à projets recherche pour le financement de thèses et de post-docs sur des sujets de recherche interdisciplinaires
- Organisation de défis et de compétitions autour de rétro-engineering de code, génération de profils et tests en masse, collectes de données,...

**4- Animation de la communauté des utilisateurs (industriels, pouvoirs publics, associations de consommateurs,...)**

- Actions de diffusion du savoir-faire et des bonnes pratiques: aide au développement d'algorithmes d'apprentissage pour les rendre transparents, par exemple pour les entreprises qui souhaitent transformer leur code existant pour être conforme à la loi et diffuser les bonnes pratiques
- Recrutement d'ingénieurs de développement logiciel qui seront amenés à interfacer les services de l'état, les industriels pour la transformation des logiciels existants

**5- Action de communication**

- Conférence annuelle scientifique, évènement de communication vers le grand public, supports du faire-savoir des résultats de la plateforme