

AVIS DU COMITE ETHIQUE ET SCIENTIFIQUE DE PARCOURSUP SUR L'ALGORITHME NATIONAL

Avis n°002-2018 du 28 novembre 2018 sur le projet d'arrêté sur le traitement automatisé ORISUP

Le CESP

Vu le règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données ;

Vu le code de l'éducation, notamment son article L.612-3 ;

Vu le code des relations entre le public et l'administration, notamment son livre III ;

Vu la loi n°78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés (loi Informatique et Libertés) ;

Vu l'arrêté du 9 mars 2018 relatif aux missions, à la composition et aux modalités de fonctionnement du comité éthique et scientifique de la Plateforme Parcoursup ;

Vu l'avis de la commission d'accès aux documents administratifs n° 20161989 du 23 juin 2016 ;

Vu la décision de la commission nationale de l'informatique et des libertés n° MED 2017-053-du 30 août 2017 mettant en demeure le ministère de l'enseignement supérieur, de la recherche et de l'innovation et la délibération n°2017-233 du bureau de la commission nationale de l'informatique et des libertés décidant de rendre publique la mise en demeure n° 2017-053 ;

Vu la délibération n°2018-119 du 22 mars 2018 portant avis sur un projet d'arrêté autorisant la mise en œuvre d'un traitement de données à caractère personnel dénommé Parcoursup ;

Vu les minutes de l'audition de Madame Claire Mathieu, chargée de mission Parcoursup par le MESRI ;

Après avoir entendu le rapport de Gerard Berry :

Considérant que, pour la première fois dans le cadre de Parcoursup, l'administration a publié- le 21 mai 2018 – le code source de l'algorithme national utilisé en vue de la pré-inscription des candidats dans les établissements d'enseignement supérieur répertoriés sur la plateforme; et ce, en application de l'article L.612-3 du code de l'éducation qui prévoit que *« la communication, en application des dispositions du code des relations entre le public et l'administration, du code source des traitements automatisés utilisés pour le fonctionnement de la plateforme mise en place dans le cadre de la procédure nationale de préinscription (...) s'accompagne de la communication du cahier des charges présenté de manière synthétique et de l'algorithme du traitement »* ; que cette disposition s'inscrit dans la ligne de l'article L. 300-2 du code des relations entre l'administration et le public sur la communication des codes sources ;

Considérant que si les règles définissant les principaux traitements algorithmiques utilisés par les administrations dans l'accomplissement de leurs missions lorsqu'ils fondent des décisions individuelles doivent être publiées en vertu de l'article L. 312-1-3 du même code, ce dont l'algorithme national est exempté en vertu de l'article L. 612-3 précité, la publication d'une description textuelle, mais scientifiquement précise de l'algorithme national Parcoursup et des programmes informatiques implémentant cet algorithme est sans précédent dans le secteur de l'éducation et de l'enseignement supérieur ;

Considérant en effet que, si la procédure d'affectation Parcoursup ne se résume nullement à un algorithme et fait une place non négligeable à l'intervention humaine et notamment à celle des commissions d'examen des vœux de chaque formation inscrite dans Parcoursup placés sous la responsabilité de chaque établissement, il importe de familiariser le public avec cet outil devenu incontournable de gestion administrative, en particulier s'agissant d'affectation de masse, comme c'est le cas pour les quelque 800 000 candidats à l'enseignement supérieur, nombre en augmentation constante ;

Considérant que si le traitement algorithmique Parcoursup comporte très peu de données à caractère personnel dites sensibles au sens de la loi Informatique et Libertés susvisée, de par sa volumétrie et l'impact d'éventuelles erreurs de programmation et fuites de données, cet algorithme doit faire l'objet d'une analyse d'impact ainsi que de strictes mesures de sécurité ;

Le CESP est d'avis que :

I. Sur la bonne compréhension de l'algorithme national

La description textuelle de l'algorithme général ainsi rendue publique est à la fois scientifiquement précise et très bien écrite ;

Les choix techniques d'implémentation fait par les concepteurs dudit algorithme sur des points non explicitement précisés par les spécifications initiales – concernant par exemple la façon d'interclasser les boursiers, les candidats non-résidents dans le ressort d'une académie ainsi que les candidats à l'internat – sont notamment très bien explicités, argumentés et illustrés dans le document publié ;

Cette description de l'algorithme est adaptée aux besoins d'analyse par des spécialistes, mais aussi par toute personne ayant une formation mathématique et informatique de niveau raisonnable dont les professeurs de lycée enseignant en informatique ;

Cependant, pour satisfaire aux exigences de compréhension par des non-spécialistes, l'administration devrait éditer une version plus grand public du document, explicable à tout le moins par des enseignants non spécialistes en informatique au niveau de l'enseignement secondaire et universitaire ; et ce, pour favoriser la compréhension des principes de fonctionnement de l'algorithme Parcoursup par tous les enseignants, les élèves et étudiants et leur famille, ainsi que tous commentateurs, y inclus les journalistes non spécialisés ;

A cet effet, Il serait opportun de réaliser par exemple un film d'animation illustrant le fonctionnement de l'algorithme, à l'instar des nombreuses vidéos disponibles sur le web à destination de divers publics et concernant des algorithmes variés.

II. Sur la validation de la programmation de l'algorithme national

L'algorithme décrit dans le document textuel ne doit être considéré que comme la spécification précise du programme informatique, exécuté sur les ordinateurs du ministère chargé de l'enseignement supérieur pour calculer les réponses faites aux candidats en fonction de leurs vœux et des classements par les commissions locales. Ce programme, écrit dans le langage de programmation Java, est beaucoup plus détaillé que l'algorithme, car il prend nombre de décisions techniques d'implémentation. Il est lui aussi consultable sur le site du ministère, et son comportement précis peut donc être comparé par tout lecteur spécialiste aux spécifications du document scientifique, pour détecter d'éventuelles erreurs d'implémentation. Une telle analyse est cependant délicate, comme pour toute réalisation informatique ;

Lorsqu'elles précisent certains points fins de l'algorithme intervenant dans la classification finale des candidats, mais non précisés dans le cahier des charges et la description textuelle de l'algorithme, les décisions techniques prises et approuvées au fil du développement par les concepteurs du cahier des charges lors de l'écriture du programme devront être explicitement à nouveau validées par ces derniers et ajoutées à la description textuelle actuelle de l'algorithme si elles n'y sont pas déjà. Cela vaut par exemple pour la prise en compte de la notion de quota de boursiers ou de candidats résidents hors académie, qui a plusieurs interprétations possibles entre lesquelles il a fallu choisir, et le traitement spécifique des « meilleurs bacheliers » ;

Pour traiter ces divers aspects, l'administration devrait donc développer des procédures de validation des programmes comparables à celles mises en œuvre dans les domaines où la qualité de l'informatique revêt un caractère critique (ex. avionique, ferroviaire) ;

Dans le cas de Parcoursup, l'algorithme étant bien défini, l'administration pourrait néanmoins aller encore plus loin dans la vérification de la conformité du programme aux spécifications de l'algorithme, en utilisant les méthodes les plus modernes pour garantir mathématiquement sa conformité du programme à ces mêmes spécifications, elles-mêmes décrites de façon mathématique ; cela correspondrait à une première mondiale dans le domaine.

III. Sur les interfaces et interactions informatiques avec les utilisateurs de la plateforme

Les utilisateurs interagissent avec la plateforme Parcoursup au moyen d'interfaces implémentées sur les navigateurs web de leurs ordinateurs, tablettes ou smartphones. Il importe donc de garantir la qualité de ces interfaces et ces interactions au même titre que celle de l'algorithme et du programme ;

L'administration devrait également mettre en place un dispositif assurant un retour des utilisateurs permettant d'identifier les points forts et les points faibles desdites interfaces et interactions ;

Elle devrait parallèlement constituer un comité d'utilisateurs – candidats, responsables de formation et développeurs de la plateforme - travaillant en commun à la qualité des interfaces et interactions homme-machine permettant la communication des candidats avec la plateforme et les formations, que ce soit par l'intermédiaire de sites Internet ou par des courriels.

IV. Sur la sécurité du système

Comme tout responsable de traitement au sens de la loi Informatique et Libertés, le ministère chargé de l'enseignement supérieur doit garantir la sécurité du système, selon les préconisations tant de la commission nationale de l'informatique et des libertés que de l'Agence nationale de sécurité des systèmes d'information ;

L'impératif de sécurisation devrait porter non seulement sur la plateforme, mais sur l'ensemble de ses données, interfaces et procédures de sauvegarde.

*