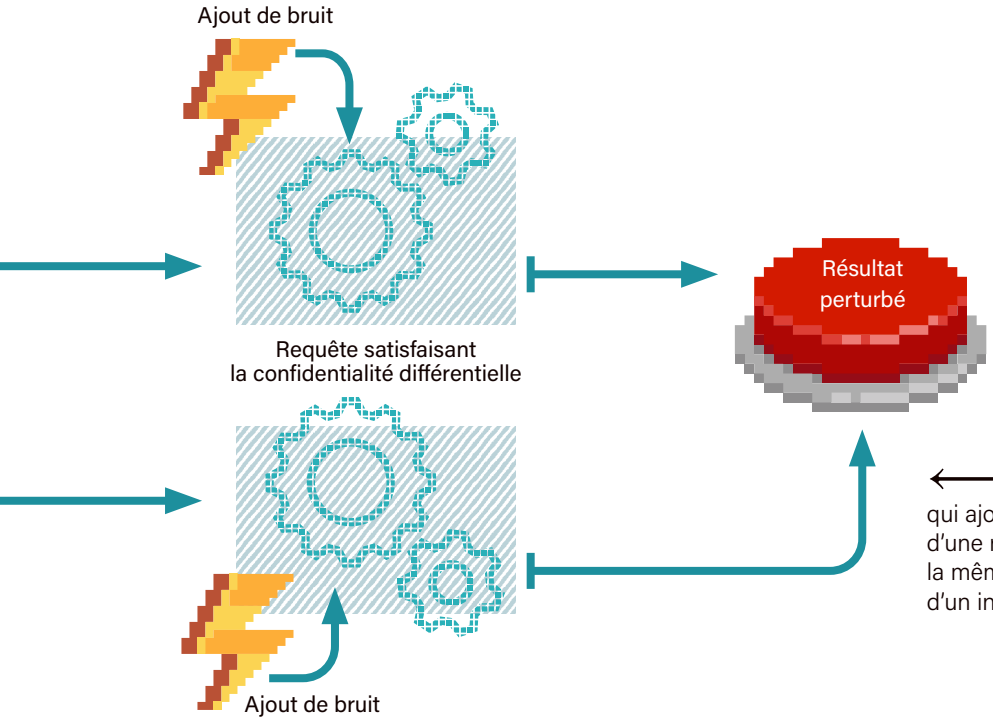

Avec la confidentialité différentielle, qui ajoute du bruit de façon aléatoire au résultat d'une requête, une analyse sera sensiblement la même avec ou sans la contribution d'un individu quelconque.

D'autres variantes encore tiennent compte de la modification de la définition de la sensibilité, à savoir la façon dont on mesure l'impact d'un individu sur la sortie de l'algorithme d'anonymisation. La définition classique considère uniquement la présence ou l'absence de l'individu, mais on peut l'étendre à la modification d'un de ses attributs. Sur un exemple concret, ce type de variation protège non pas toutes les visites individuelles d'un site web par un individu, mais le fait qu'il ait déjà visité ou non le site. Enfin, si la confidentialité différentielle est typiquement adaptée à des données tabulaires (en tableau), elle peut se décliner à d'autres types de données à protéger, par exemple des données de type graphe dans le cadre d'un réseau social où les données représentent les individus et les liens qui les unissent.

UNE BIBLIOTHÈQUE DE BRIQUES

L'univers des algorithmes satisfaisant la confidentialité différentielle a en parallèle lui aussi fortement crû. Ces algorithmes dépendent du type de données à anonymiser (données sociodémographiques dans un tableau, séries temporelles, ensembles de nœuds et d'arêtes d'un réseau social…), de la catégorie d'informations à publier, du contexte de calcul (centralisé ou distribué). Malgré le large éventail de solutions disponibles, lorsqu'on ne trouve pas chaussure à son pied, il est nécessaire de concevoir un nouvel algorithme dont les garanties de confidentialité différentielle sont à prouver formellement. Cette tâche est loin d'être triviale, mais des outils ont été proposés. D'une part, certains offrent un cadre de démonstration clair listant les conditions formelles nécessaires pour qu'un algorithme satisfasse la confidentialité différentielle. À charge pour le concepteur de démontrer que son algorithme les remplit. D'autre part, des outils de vérification automatique établissent une preuve formelle, ou non, de la confidentialité différentielle.

Tous ces efforts sont grandement facilités quand l'algorithme en question peut être décomposé en briques élémentaires pour lesquelles la confidentialité différentielle est déjà assurée. Le concepteur pioche alors ce dont il a besoin dans des bibliothèques existantes. Pour ce faire, un écosystème d'outils a été développé, faisant le pont entre les travaux académiques et ceux de l'industrie. La bibliothèque Ektelo propose un ensemble générique de briques de bases pour développer des algorithmes obéissant à la confidentialité différentielle. IBM, Google et Microsoft distribuent aussi leurs propres bibliothèques. En 2020, Google a par ailleurs donné accès à une version de la bibliothèque d'apprentissage machine TensorFlow incorporant la confidentialité différentielle afin d'entraîner des modèles respectueux de la vie privée. Ces bibliothèques assurent la gestion du budget de vie privée, refusant l'exécution d'une requête si ce dernier est épuisé.