

De plus en plus de données sont créées à chaque instant par nos activités numériques : traces de navigation, achats en ligne, images de vidéosurveillance, etc. Regroupés sous le vocable « Big Data », des projets innovants utilisent ces données et annoncent un monde parfois meilleur, parfois effrayant. Ces données massives sont notamment le carburant des techniques les plus spectaculaires d'intelligence artificielle et de robotique.

**?** En quoi le Big Data va-t-il transformer nos vies : quelles sont les opportunités et les risques ?

**e** SNT2043

## Pour aller plus loin

- ▶ « Les enjeux du Big Data », Arte, Futuremag
- 📖 « Big data : le boom des données numériques », Futura Sciences
- ▶ « Black Mirror, quand la technologie vire au cauchemar », France Culture

# BIG DATA : LES DONNÉES

## 1 Un monde de données

Le terme Big Data, « données massives », désigne des volumes de données tellement importants qu'ils ne peuvent plus être traités avec des outils de gestion classiques. Le Big Data est l'un des grands enjeux informatiques de notre époque, notamment pour la recherche et le développement. Il soulève également des questions en termes d'éthique et de libertés individuelles. Comment utiliser ces milliards de données au bénéfice de tous ?



▲ Le Big Data vise à extraire l'information pertinente des masses de données collectées.

**e** SNT2041

▶ À quoi servent les Big Data ?

## 2 Le deep learning

Le deep learning, apparu en 2006, est une technique d'apprentissage automatique non supervisé. Elle nécessite de disposer d'un très grand nombre d'exemples déjà connus pour entraîner le système à reconnaître de nouveaux éléments.

Dans bien des cas, l'usage des techniques d'apprentissage non supervisé est lié à un autre changement majeur des dernières années : l'accès à des masses de données, via Internet et les réseaux sociaux. C'est pourquoi machine learning, deep learning et big data sont étroitement liés. Un autre facteur clé est la démocratisation de la puissance de calcul, notamment grâce aux architectures parallèles utilisant des processeurs graphiques (GPU), capables de faire tourner des algorithmes de machine learning en des temps raisonnables.

« Deep learning, big data, réseaux de neurones... pourquoi l'intelligence artificielle revient-elle maintenant », par Thierry Lucas, publié le 12 mai 2015, L'Usine Digitale.

## 3 Big Data : big brother ?

Le document de planification du « Conseil de l'État des élites chinoises » soutient à ce titre le projet gouvernemental d'une gigantesque base de données permettant la mise en place d'un système de « crédit social » d'ici à 2020. Ce programme national pourrait inclure les notes et les données récoltées sur les sites marchands, tels qu'Alipay Pay avec le programme « Sesame Credit » (développé par Ant Financial Services Group, filiale du groupe chinois Alibaba). Il

Le Big Data permettrait de ficher toute une population au détriment de la vie privée.

est stipulé dans ce document que « la base de données de crédit social forgera un environnement glorieux de l'opinion publique basé sur le maintien de la confiance » et « récompensera ceux qui signalent des actes d'abus de confiance ». Le concept de « notation sociale », déjà mis en place sur les sites marchands et les banques, est depuis 2014 testé sur le plan politique par les autorités locales dans une vingtaine de villes. C'est ainsi qu'à Shenzhen, des écrans



géants placés aux carrefours des rues affichent les visages des piétons qui ne respectent pas les feux de signalisation. Pour cette infraction mineure par exemple, 20 points sont enlevés aux citoyens de la ville de Suqian, sur les 1 000 points crédités sur leur compte.

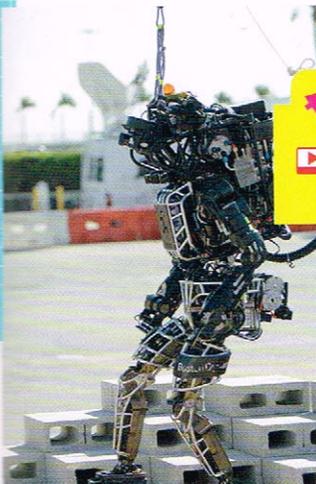
« Cybercontrôle en Chine : l'omni-surveillance à l'ère du numérique », par A. Bayen, publié le 21 déc. 2018 © Ina - La Revue des médias.

## 4 Apprendre aux robots

La robotique a bénéficié des avancées en intelligence artificielle dues au Big Data et à leur utilisation par le *deep learning*. La possibilité d'apprendre de nouveaux comportements à partir d'exemples existants permet d'atteindre une efficacité inégalée. Mais pour qu'un tel système fonctionne correctement, il faut disposer de données fiables et représentatives des situations que le robot rencontrera pendant son utilisation. C'est ce qui lui permettra d'interpréter son environnement et d'agir en conséquence.

## 5 La médecine de demain

Le *deep learning* désigne les algorithmes capables de reconnaître des schémas remarquables parmi un très grand nombre de données disponibles. Il permet déjà d'accélérer l'identification des cancers à partir des résultats des scanners et des IRM en utilisant les techniques pour la reconnaissance d'image. La collaboration entre le médecin et la machine promet un diagnostic encore plus rapide et plus précis, augmentant ainsi les chances de guérison. Des études sont en cours pour utiliser le Big Data en vue de proposer les traitements les plus adaptés aux patients. Par ailleurs, le traitement des données pourra être utilisé dans la sélection des personnes pour les essais cliniques et l'analyse de leurs résultats afin de concevoir des médicaments plus efficaces.



**e** SNT2042  
Le robot Atlas de Boston dynamics