

L'éthique et les Big Data en Santé : *entre promesses et garde-fous éthiques*



Pour toute utilisation du contenu de cette présentation, veuillez citer l'auteur, son organisme d'appartenance, le titre et la date du document, ainsi que le volet 1 de l'atelier 2020 « Usages éthiques des Big Data en biosciences » de la Plateforme « Ethique et Biosciences » de Toulouse. Merci.

La Datafication du monde

LA DATA-SPHÈRE

Toutes les minutes :

- 350 000 Tweets
- 15 millions de SMS
- 200 millions de mails
- 250 gigaoctets d'information sont archivés sur Facebook
- 1 740 000 gigaoctets (Go) d'informations sont publiés dans le monde

Tous les jours

- Google traite plus de 24 peta-octets de données soit 24 millions de milliards d'octets
- En 2013, 1,01 Md€ d'objets connectés
- En 2020, 100 Milliards d'objets connectés



La numér-ère

La Digitalisation de la société

Election du Pape



Concert de musique



Le changement de paradigme



Infosphère : Individuel / Temps réel

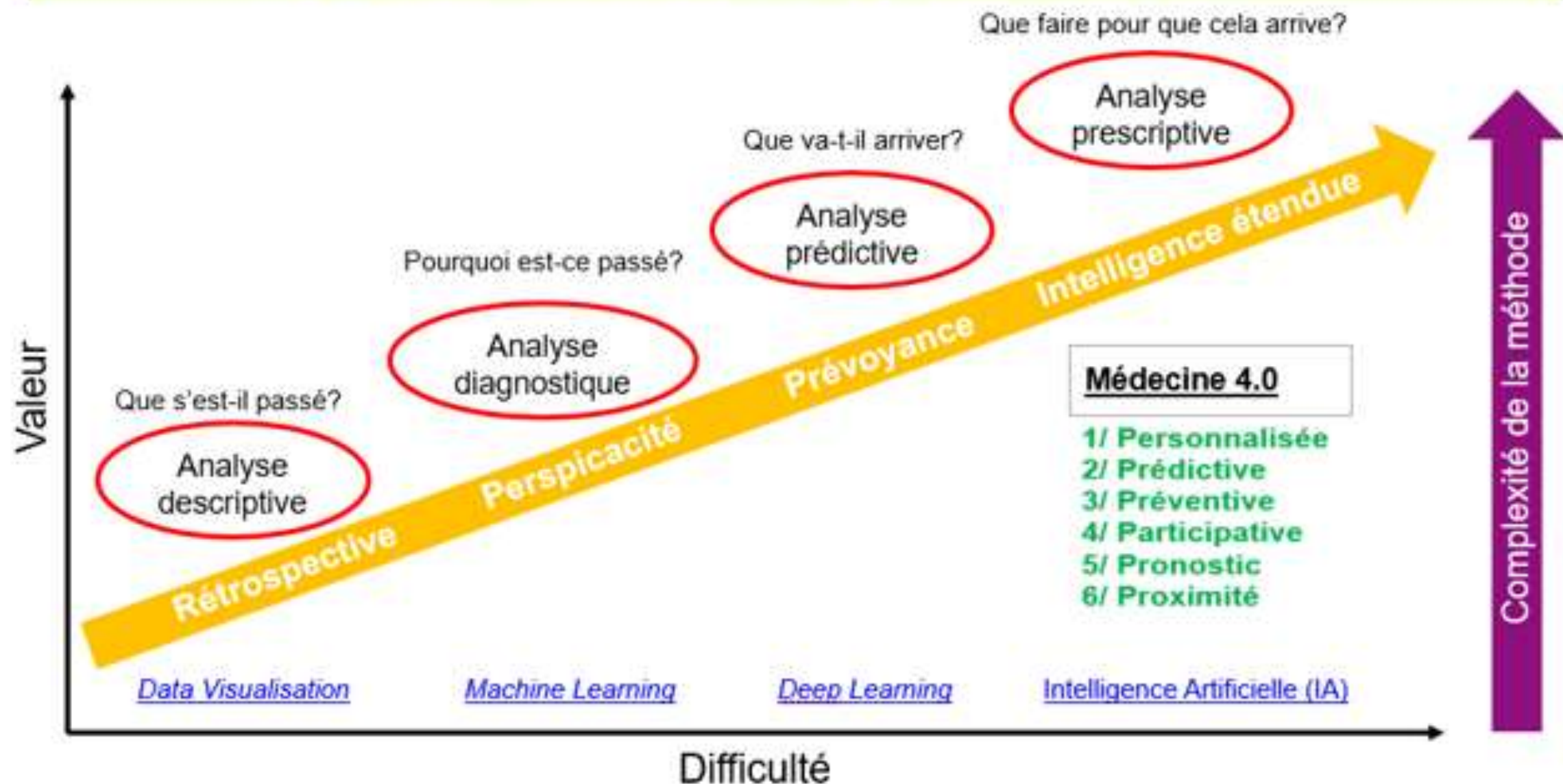
- Dématérialisation des relations
- Désintermédiation
- Economie collaborative

Caractéristiques des *Big Data* (7V) :

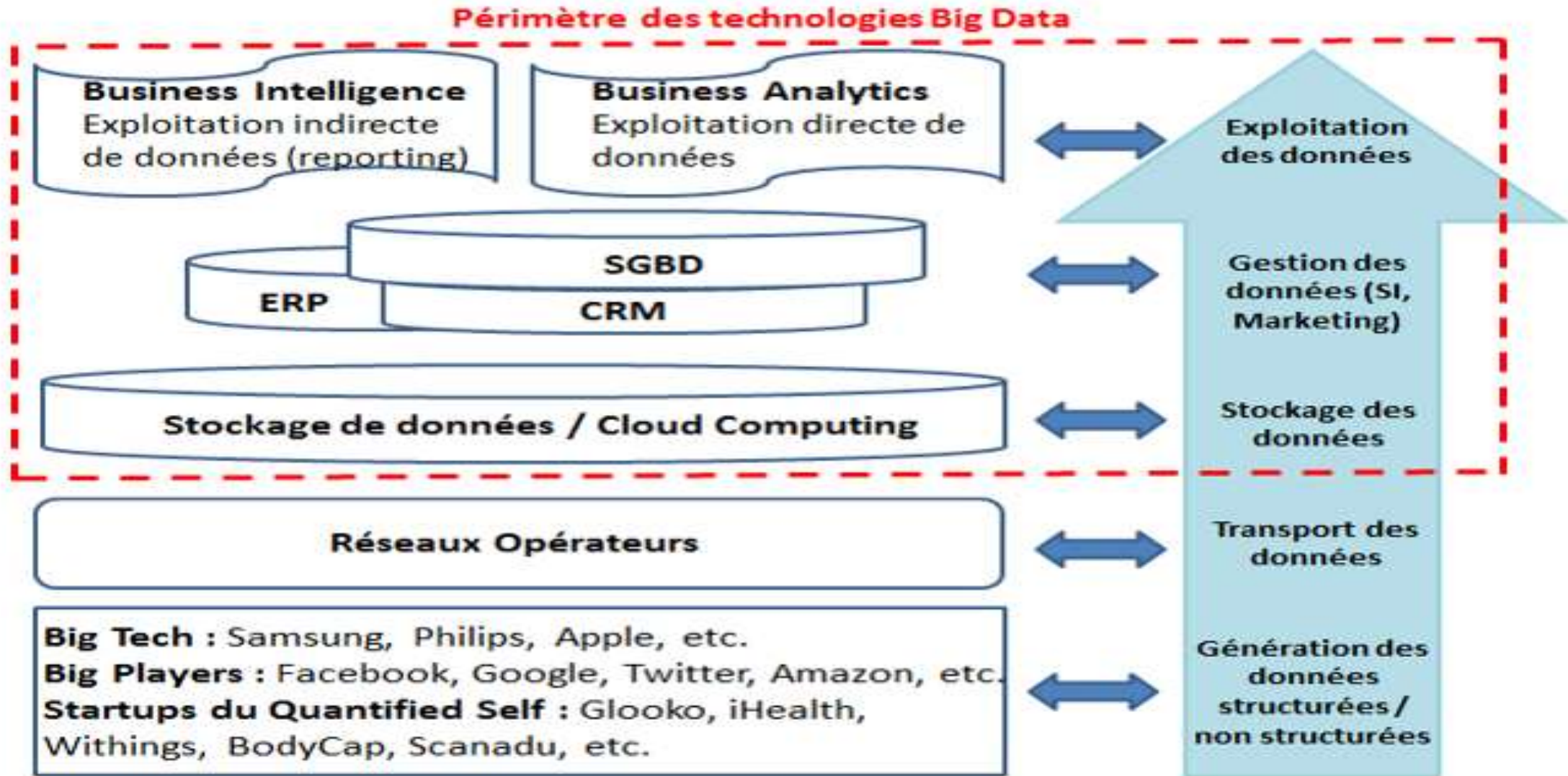
- 1/ Volume
- 2/ Variété
- 3/ Vitesse
- 4/ Véracité
- 5/ Volatilité
- 6/ Visualisation
- 7/ Valeur

**Agrégation
Algorithmes auto-
apprenants**

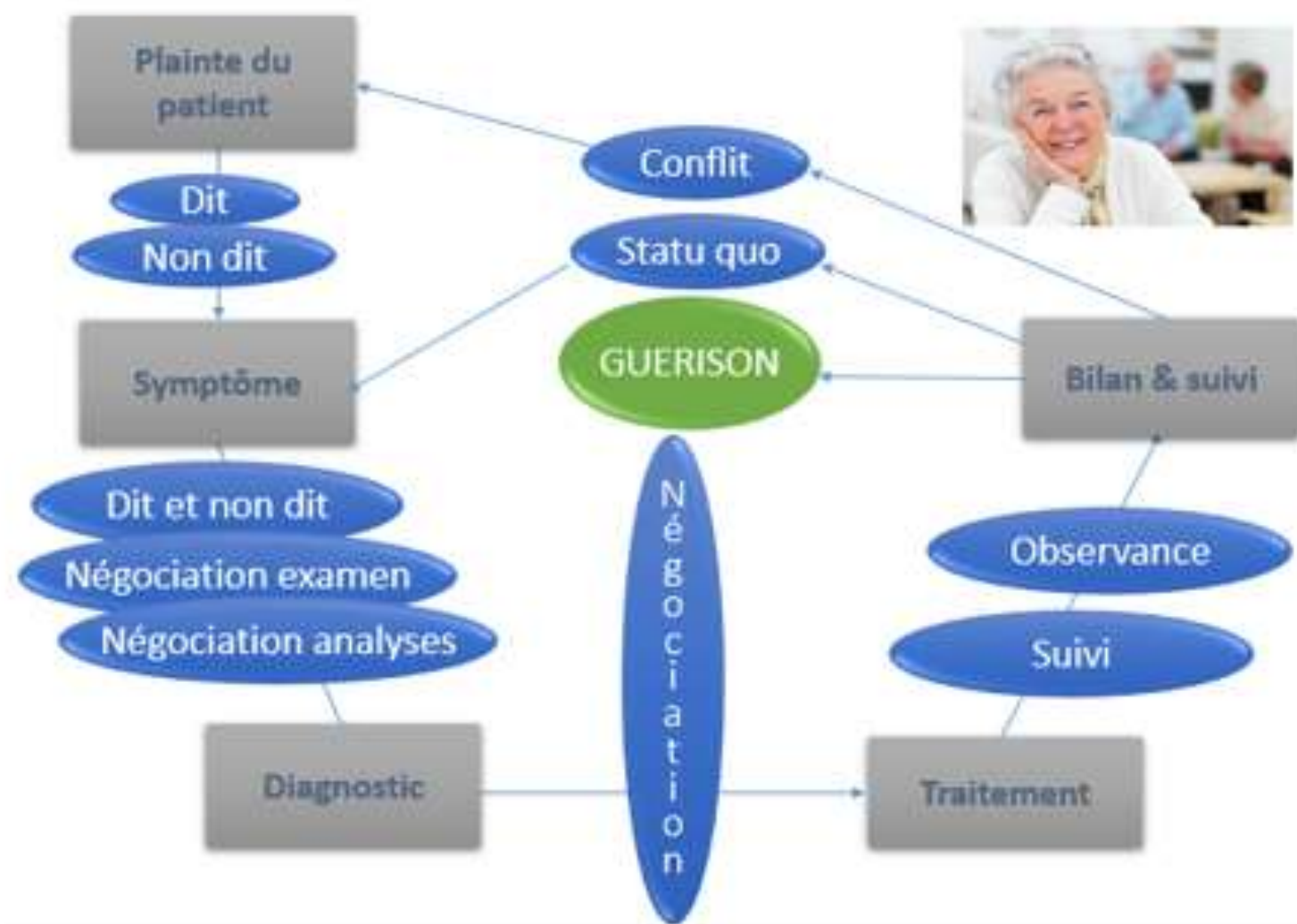
L'Evolution de la Data Science médicale



Périmètre des technologies et acteur des Big Data



La relation Médecin-Patient



Où exactement?

Intensité ?
Saignement ?

A quelle heure?

Ecoulement?

Depuis quand ?

Maladies
Ca brûle
Nébul brutal ?

Ca dépend

Ca va ça vient
Ménopause

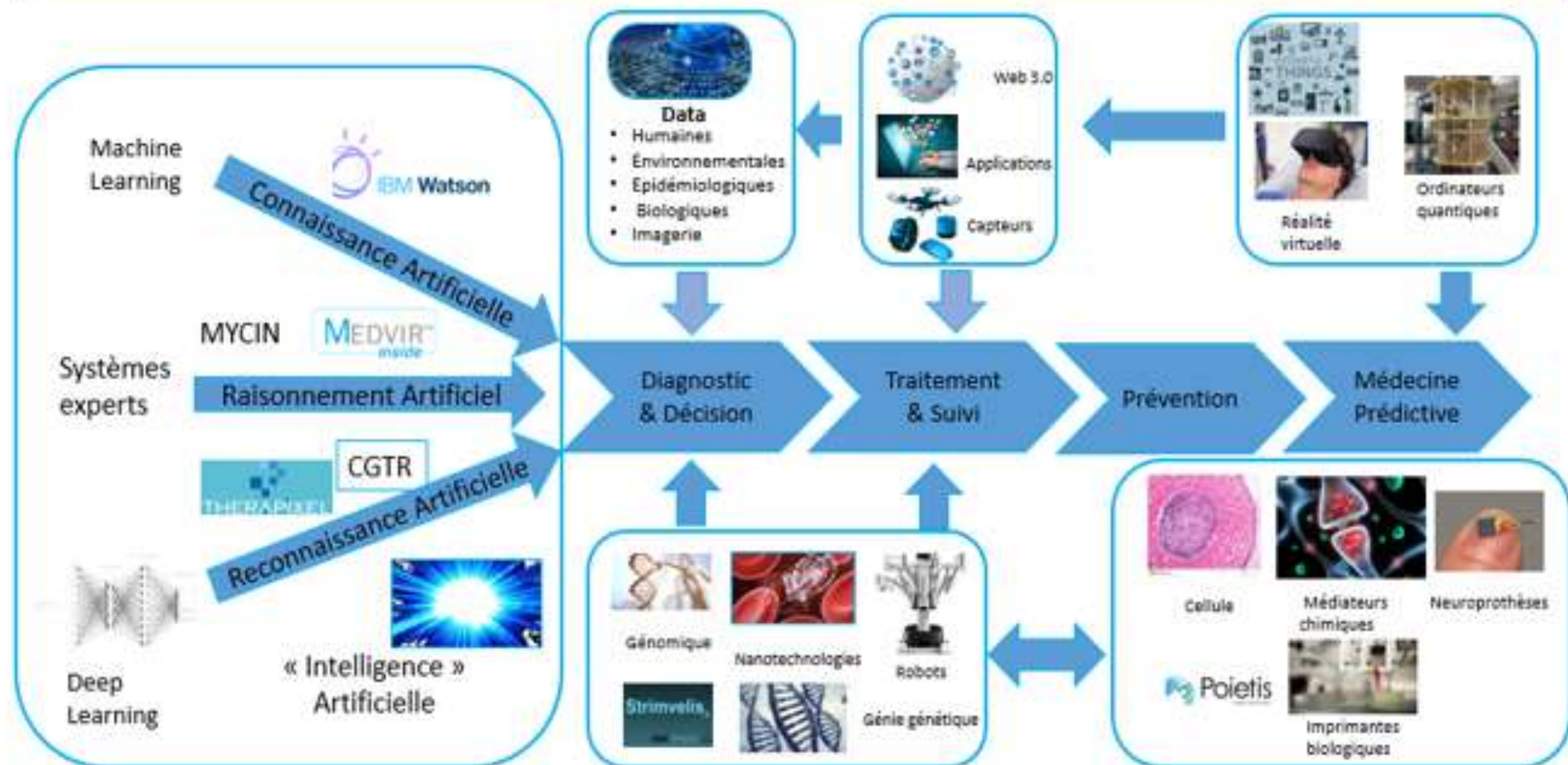
Tout le temps

Transpiration

Ca fait longtemps

Maladies
Ca
Tous les
gratte

La Médecine 4.0



Les aspects multidimensionnels de l'IA

Définition de l'IA (faible)

- Simuler l'intelligence humaine en reproduisant certaines de ses fonctions cognitives
- Reproduire les fonctions sensorielles des êtres sensibles
- Acquérir des connaissances et apprendre à partir des données
- S'adapter à de nouvelles solutions

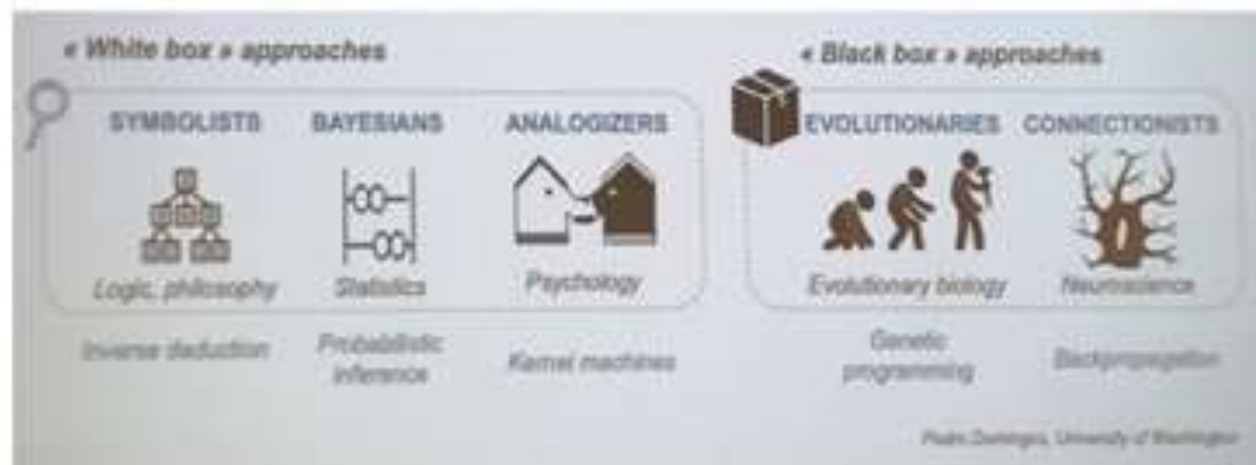


Applications pour l'IA

- Reconnaissance visuelle massive
- Prédiction
- Support à la décision
- Allocation des ressources
- Amélioration productivité / sécurité



IA faible Vs IA forte



Intelligence artificielle



IA faible

- ✓ Apprentissage
- ! Compréhension
- ! Imagination
- ! Conscience de soi

- ✓
- ✓
- ✓
- ✓

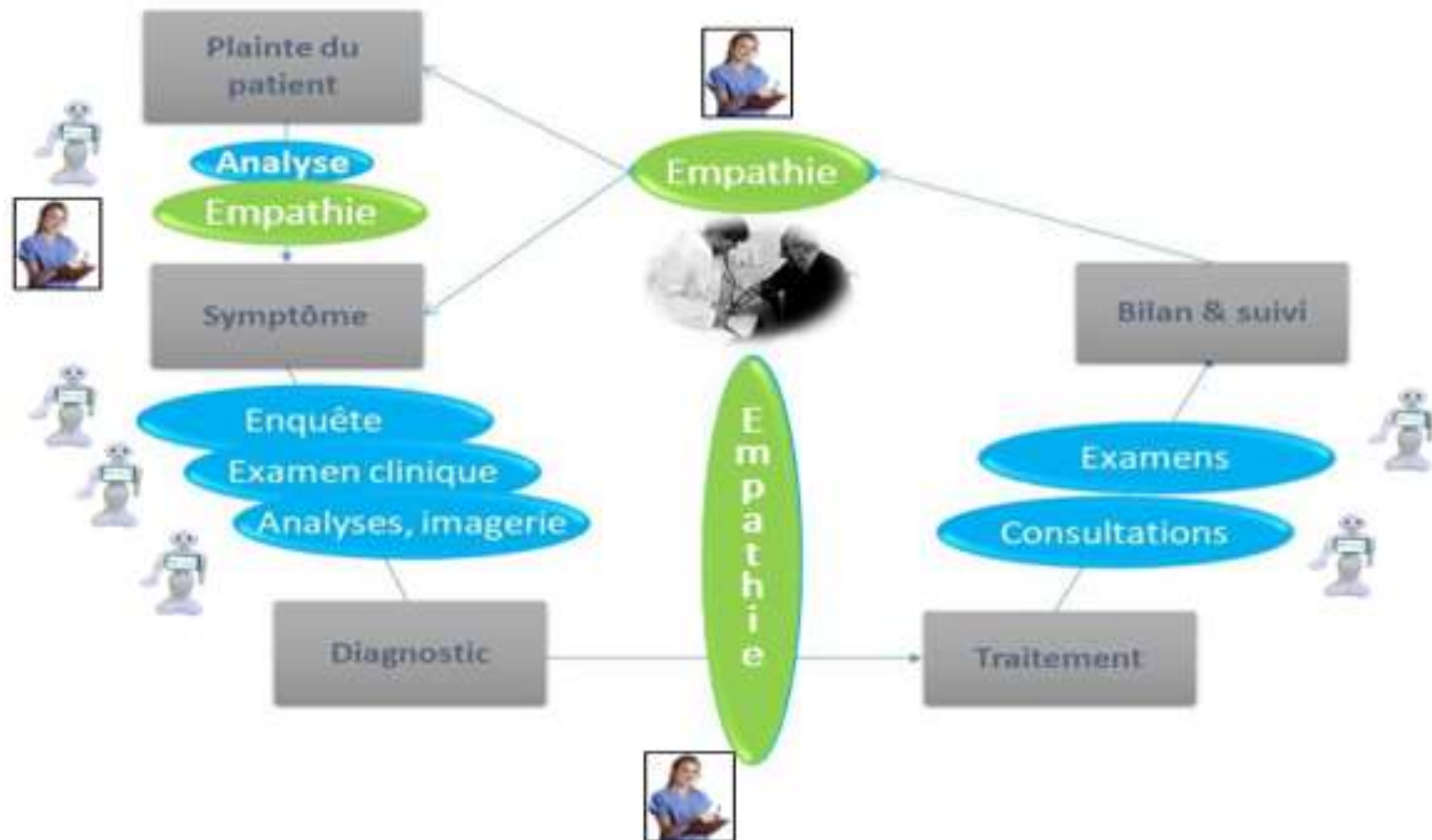


Intelligence humaine

IA forte

Il n'existe au monde aucun système d'IA forte

Le Médecin et la Machine



La Numérisation de la Médecine



Efficiences des parcours de soins et des organisations

Sécurité sanitaire
(bon usage, pharmacovigilance des produits de santé, indication par indication)

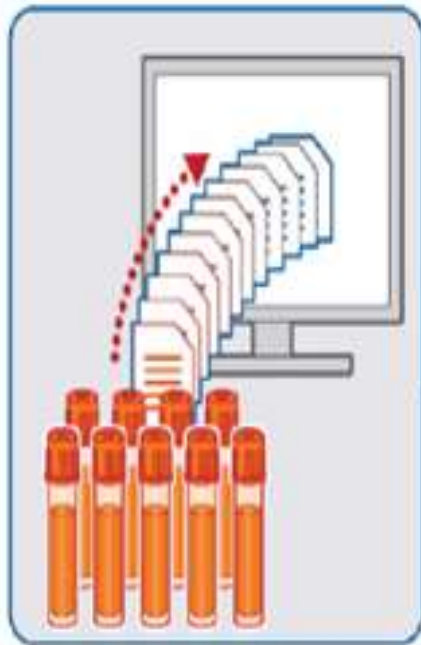


Prise en charge individuelle des patients
(médecine personnalisée)

Prise en charge collective
(santé populationnelle)

Le radiomique

L'IA permet d'exploiter des données brutes d'imagerie médicale pour en extraire de manière automatique des informations d'ordre biologique et moléculaire. Intérêt : Evite de réaliser une biopsie au patient.



APPRENTISSAGE

L'algorithme est nourri de centaines d'images de tumeurs et de leurs données biologiques. Il apprend à associer les caractéristiques d'un cliché à la composition cellulaire du cancer.



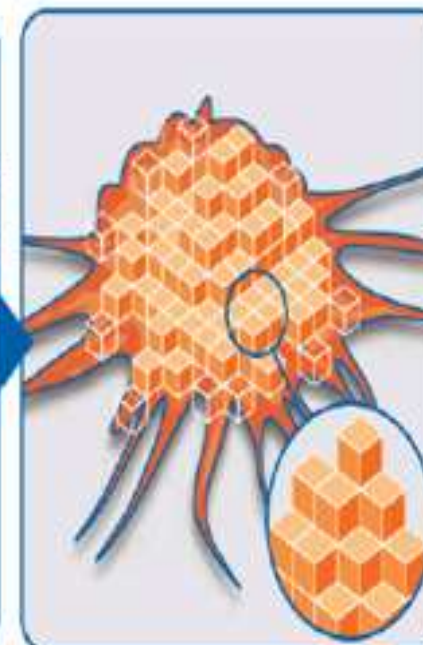
ACQUISITION D'IMAGE

En phase de diagnostic, le patient passe un scanner dont l'image sera décortiquée et analysée par le système d'intelligence artificielle entraîné.



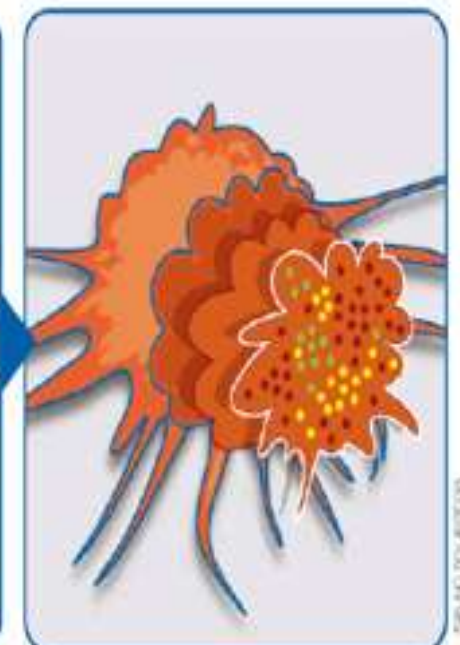
ANALYSE ALGORITHMIQUE

L'algorithme localise la tumeur et en extrait différentes caractéristiques : forme, volume, surface.



VOXELS

La tumeur est détaillée en voxels de 1 mm^3 (l'équivalent en 3D des pixels). L'IA détermine les relations spatiales entre les voxels et différents niveaux de couleurs.



DIAGNOSTIC BIOLOGIQUE 3D

En exploitant ces données, l'algorithme parvient à déterminer des « biomarqueurs radiomiques » comme la concentration en cellules immunitaires (lymphocyte T) de la tumeur.

Préoccupations des citoyens face à la numérisation de la société

81% des français préoccupés par la protection de leurs données personnelles

85% des Français très préoccupés par la protection de leurs données personnelles sur Internet

91% des Français préféreraient stocker leurs données chez eux et non via un service en ligne



84% des internautes français âgés de 15 à 64 ans se disent inquiets de l'usage fait de leur *Data*

Sources: Etude de l'Institut GSA pour Orange : « Les Français & la protection des données personnelles » - Février 2014 / Sondage commandé par Lima - Juin 2015 / Etude pour Havas Media réalisée par l'institut Toluna et publiée - Jeudi 25 septembre.

Interrogations autour des systèmes algorithmiques

Valeur et pertinence scientifique

Finalités de la collecte et du traitement

Formation de l'utilisateur

Gestion du cycle de vie

Obtention du consentement



Sourcing issus de la fusion de données

Respect de la dimension humaine

Jugement du professionnel

L'interopérabilité, stockage et sécurité

Discrimination et catégorisation

Attentes des professionnels de Santé

Intégration adaptée
dans les pratiques
quotidiennes

Formation et
accompagnement

Fluidité
Professionnel -
Patient

Sécurité des
données récoltées



Connaissance et
accès au patrimoine
de données de santé

Cadre juridique
économique adapté

Interopérabilité

Fiabilité des outils IT

Clarification de
la gouvernance

Valeur intrinsèque de la donnée (1/2)

Caractère de pertinence	Caractère opérationnel	Caractère consolidable	Caractère nominal	Caractère relationnel
Fidèle	Facile à établir (simplicité)	Quantifiable	N'est jamais unique	Reconnu / Acceptable
Juste et stable	Facile à utiliser	Cumulable	En nombre limité	Motivant / Fédérateur / Mobilisateur
Précis et sensible	Communicant			
Utile	Utile à la prévention			

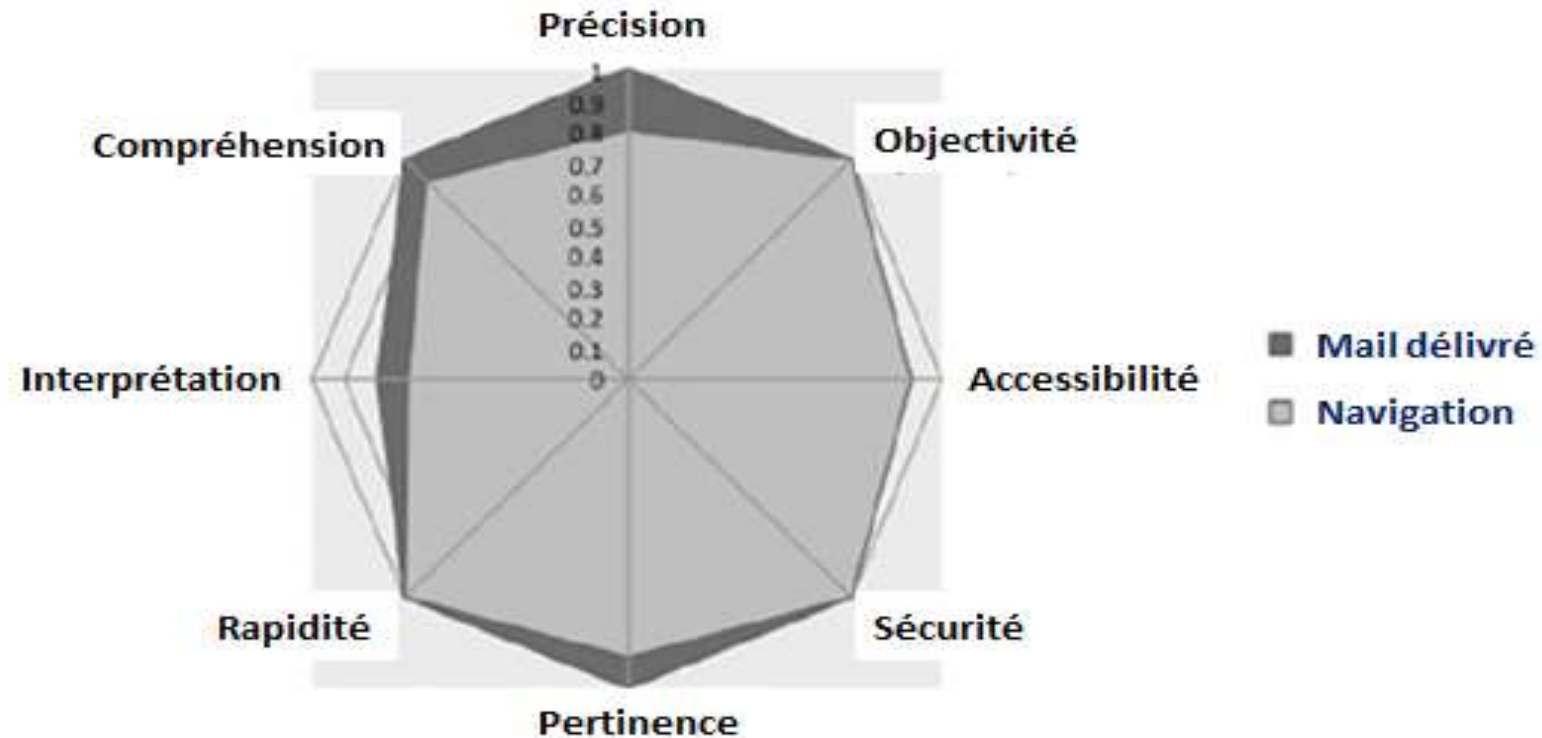
Valeur intrinsèque de la donnée (2/2)

Infosphère		
Domaine d'étude	Stratégique et méthodologique	Structural et technologique
<u>Epistémologie :</u> Donnée / Information	Applicabilité Richesse Adaptabilité Ordonné Flexibilité Evolutibilité Performance Réutilisabilité Pragmaticité Consolidation Fonctionnalité / Opérationnalité	Cohérence / Sens Exactitude Intégrité Exhaustivité Authenticité Fiabilité Robustesse Légitimité Nominal / Base de données
Domaine d'étude	Organisationnel et réglementaire	Relationnel et culturel
<u>Epistémologie :</u> Donnée / Information	Pertinence Stabilité / Continuité Assurance Pluridisciplinarité Systématicité Référence Normativité Maintenabilité Réglementation Auditabilité Coordination	Sécurité Confidentialité Fédération Diffusion Accessibilité Convivialité Universalité Disponibilité Coopération

Qualité de l'information

- ✓ La qualité ne constitue pas une notion objective unique ;
- ✓ Elle se définit comme l'assurance que le contenu de l'information n'a pas été modifié au cours de sa transmission ;
- ✓ Elle conditionne la réalité du consentement libre et éclairé de l'assuré vis-à-vis du professionnel ;
- ✓ La qualité d'une information peut se caractériser en cinq éléments et critères bien déterminés :
 - L'intégrité ;
 - L'exactitude ;
 - La précision ;
 - La validité ;
 - L'authenticité.

Indicateurs de qualité de la donnée



Catégories de la QI	Dimensions de la QI
Intrinsèque	Précision, Objectivité, Crédibilité
Accessibilité	Accès, Sécurité
Contexte	Pertinence, Valeur ajoutée, Complétude, Exhaustivité, Quantité de données
Représentation	Interprétation, Facilité de compréhension, Représentation concise, Représentation cohérente

Grands principes (Cf. Loi Informatiques et Libertés - 1978)

5 grands principes ressortent de la **Loi Informatique et Libertés de 1978** :



Art.
6, 2°

1. **Le principe de finalité** qui suppose que la collecte de données personnelles correspond à un besoin bien défini par celui qui opère le traitement, et que l'objet du traitement soit légitime.



Art.
6, 3°

2. **Le principe de proportionnalité et de pertinence** quant à lui a trait à la collecte des informations : on ne collecte que les informations nécessaires au traitement; les données doivent être adéquates, pertinentes et non excessives au regard des objectifs poursuivis.



Art.
6, 5°

3. **Le principe de conservation limitée des données** qui requiert de déterminer une durée de conservation en fonction de la finalité du traitement



Art.32,
38
39, 40

4. **Le principe de la garantie des droits essentiels de chacun** quant à l'accès aux données le concernant : droit d'information, droit d'accès, droit de rectification, et droit d'opposition.



Art.
34

5. **Le principe de sécurité et de confidentialité des données** qui réserve l'accès aux fichiers de données personnelles aux seules personnes autorisées et concernées.

De nouvelles obligations pour le responsable du traitement (Cf. RGPD)



Privacy by design & by default

- Protection dès la conception (pseudonymisation, minimisation)
- Protection par défaut



Analyses d'impact

- Avant mise en œuvre
- Évaluation des risques
- Liste des traitements concernés à préciser (Cnil)



Accountability

- Documentation
- Preuve du respect du règlement



Sécurité

- Pseudonymisation et cryptage
- Évaluations régulières
- Adhésion Code de conduite



DPO

- Nouveau CIL
- Interne ou externe
- Rôle : information, conseil, contrôle

Valeur d'exploitation de la donnée

- Bienfaisance : Diffusion appropriée de la connaissance médicale envers l'utilisateur de SI constitue un bien-fondé et une légitimité d'action ;
- Autonomie : Information préalable (claire, précise, adaptée et compréhensible) garantit le consentement éclairé de la personne. Ce dernier agit librement et réalise un acte autonome : intentionnel, volontaire et indépendant ;
- Non-malfaisance : Accès limité aux données médicales personnelles selon le profil et la nature de l'utilisateur améliore donc la sécurité, la confidentialité et la protection de ces données

Ethique au cœur de la révolution digitale

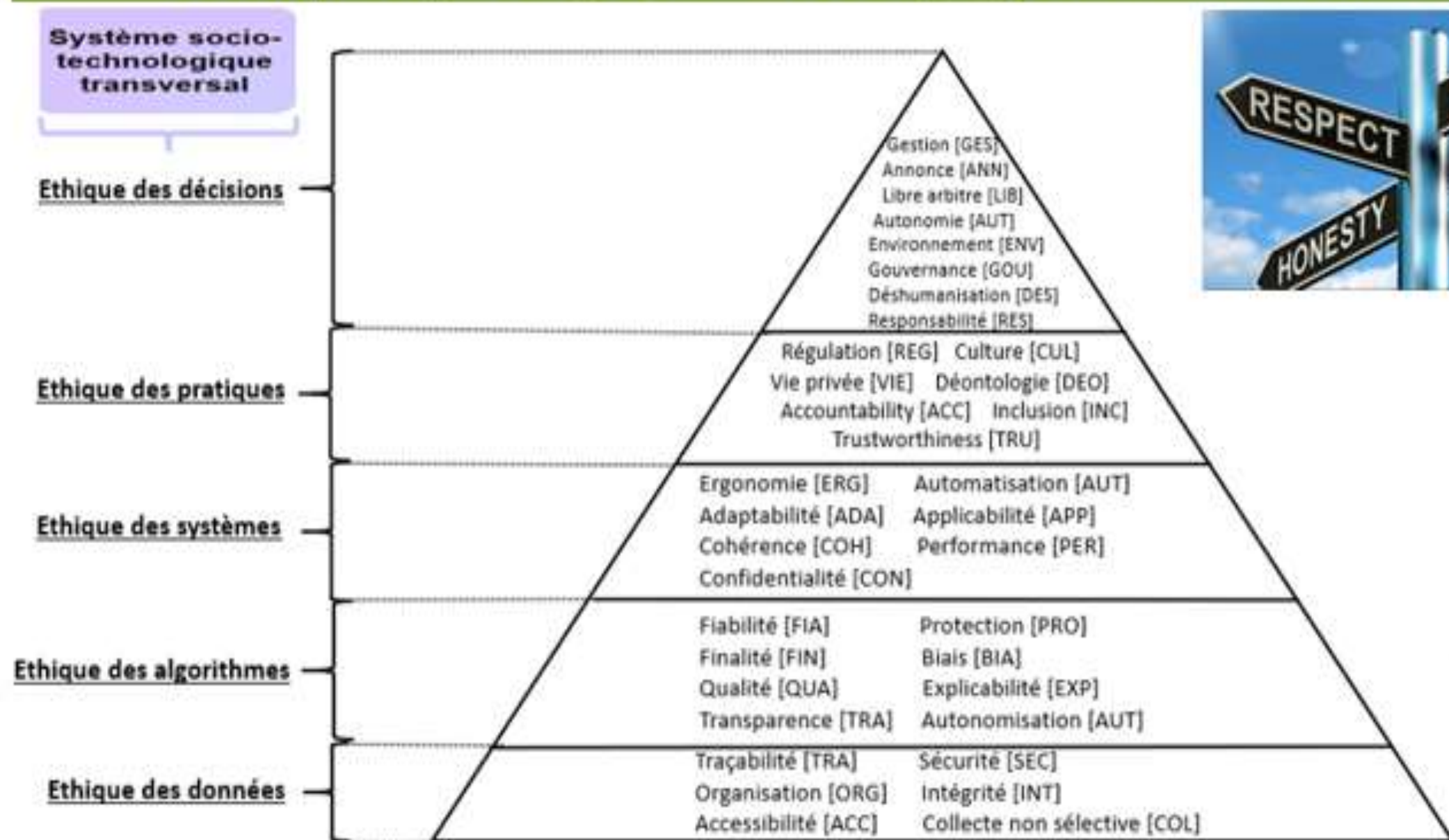
- ✓ Donner du sens à une action (Cf. Responsabilité)
- ✓ Conflit de devoirs et des droits (Cf. Déontologie)
- ✓ Ensemble des règles de bonnes pratiques
- ✓ Questionnement sur les valeurs humaines et morales qui sous-tendent l'action



Systeme socio-
technologique
transversal

Ethique des données – Ethique des algorithmes – Ethique des systemes – Ethique des pratiques (gouvernance et gestion des données) – Ethique des décisions

L'Éthique by Design pour accompagner la Médecine 4.0



Data éthique (1/2)

1. La data doit être définie précisément en rapport avec son ontologie⁶, c'est-à-dire dans le cadre du métier que l'on souhaite modéliser.
2. Elle doit être référencée dans un dictionnaire de données accessible et à jour, incluant son nom, son type, ses caractéristiques et sa définition exacte.
3. Elle doit être juste. Dans le cas d'une incertitude, celle-ci doit alors être connue et enregistrée avec la donnée.
4. Sa date et son heure précises de collecte doivent être connues et enregistrées.
5. Son mode de collecte, intégrant les différentes sources possibles pour cette donnée (par exemple un questionnaire collecté par téléphone et par internet), doit également être renseigné.

Data éthique (2/2)

6. La donnée doit être présente ou être explicitement déclarée comme manquante.
7. Elle doit être cohérente, c'est-à-dire qu'elle varie dans les limites définies dans le dictionnaire. De même, elle ne doit pas non plus être en contradiction avec une autre valeur liée à la même observation.
8. Elle doit être unique, à savoir qu'une observation ne doit pas donner naissance à deux entrées dans la même entité.
9. Elle doit être conforme, licite et validée, à savoir qu'elle respecte les règlements et standards de gouvernance internes ainsi que les règlements externes en vigueur (par exemple RGPD).
10. Elle doit être utile et de valeur : on ne stocke pas dans un SI des données sans avoir *a minima* un objectif d'utilisation ou de valorisation envisagé.

La loyauté d'un algorithme

Explicabilité et Transparence

(Décret / Loi République Numérique)

Qualité de prévision et de justesse de décisions

(Boite noire / Mode de fonctionnement opaque)

Biais, Jugement de valeurs et Discrimination

(Individuel ou Collectif / Intentionnel (Testing))

Degré d'autonomie et d'encadrement

(Responsabilité / Régulation)

Les préconisations tournées vers la Médecine 4.0

Le médecin généraliste va devoir plus s'intégrer aux organisations et aux pratiques de télémédecine

Téléconsultation au sein des cabinets, télé-expertise, télésurveillance de la compliance des malades chroniques,...

Le médecin généraliste va devenir l'un des acteurs clé de la santé connectée

En améliorant le lien ville-hôpital, la coordination, la proximité, et l'éducation thérapeutique.

Recommander des applis santé et objets connectés.

Réalisation de réunions professionnelles sur l'IA pour les acteurs de la Médecine

Recherches sur une médecine transdisciplinaire, de précision et personnalisée

Organisation centrée sur les NTIC afin de faciliter le parcours patient

Personnalisation de la relation Médecin - Patient et développement des services de proximité.

Renforcer la formation (initiale et / ou continue) des médecins dans le Numérique et l'Ethique

Le lien humain et le contrat de confiance Médecin-Patient doit être renforcé avec l'émergence de l'IA.

Faire évoluer le Code de déontologie médicale en prenant en compte le contexte digital

Plus de 60% des articles du Code sont impactés directement par le numérique.

SWOT sur les Big Data (1/2)

Forces	Faiblesses
Développement d'outils analytiques efficaces au service de la santé	Lente conduite du changement organisationnel
Emergence de nouveaux services pour la prise en charge des soins	Pénurie de compétences pour gérer les NTIC
Renforcement de la compétitivité entre les industriels de la santé	Equilibre instable entre intérêt collectif et protection des personnes
Analyse prédictive <i>via</i> les algorithmes de traitement	Fiabilité et intégrité des données exploitées
Elaboration d'environnements analytiques	Maîtrise et traçabilité des données massives
Création d'une version unique et réutilisable de la réalité	Obtention du consentement libre et éclairé du propriétaire de la donnée
Compréhension et anticipation de comportements et de parcours de l'utilisateur de santé	Conservation limitée des données
Personnalisation de la médecine (segmentation et ciblage plus fins)	Contrôle de l'accès des données
Médecine à la fois participative, préventive et prédictive (Médecine 4.0)	Sensibilisation et formation à une « culture de la donnée »
Automatisation des processus de prise en charge de la santé	Changement du paradigme médical

SWOT sur les Big Data (2/2)

Opportunités	Menaces
Généralisation du Cloud Computing	Encadrement réglementaire flou et régulation limitée
Appels à des projets publics de recherche (e-santé, m-Health, etc.)	Usages internes et externes non éthiques des Big Data médicales
Réduction des dépenses de santé	Non-respect des droits du citoyen (droit à l'information, à l'oubli, d'opposition, d'accès et de rectification)
Optimisation des parcours de l'utilisateur de santé	Déresponsabilisation des professionnels de la santé
Création de cohorte virtuelle de patients notamment en Recherche clinique	Discrimination, catégorisation et individualisation de la société
Détection rapide des signaux faibles lors d'épidémie ou d'effets indésirables graves	Sécurité et protection de l'information médicale
Développement de l'auto-surveillance médicale (outil de gestion)	Création d'hypercondriaques <i>via</i> le Quantified Self (m-Health)
Réalisation de solutions d'aide au diagnostic	Non-respect de la vie privée et des libertés

Règles de gouvernance des Big Data

- Politique de management autour de la donnée de santé
- Méthodes de gouvernance aux données prioritaires
- Sourcing et modes d'exploitation de cette donnée
- Fiabilité et la qualité d'une information
- Responsables protection de l'information
- Comités de pilotage et d'éthique
- Procédures internes et des guides de bonnes pratiques
- Dimension humaine (comportements et éducation) et organisationnelle
- Stratégie de réduction de la donnée numérique
- Ne pas conserver l'information plus longtemps que nécessaire

Les défis de la Médecine 4.0

Restituer

Anticiper

Certifier

Fédérer

Comprendre



Servir

Valoriser

Coordonner

Maîtriser

Gagner

Imaginer

Former

Conclusion

- La révolution digitale en Médecine est inéluctable! Il est essentiel de bien l'accompagner et l'encadrer.
- Evolution des métiers : les missions ne changent pas, contrairement aux fonctions et aux activités.
- Devenir un acteur clé de la santé connecté : Service de proximité hyper-personnalisé & relais (Triptyque : simplicité, utilité et hybridation).
- Un algorithme sans conscience éthique est comme une personne sans âme, leurs existences n'ont pas de sens pour l'humanité « Le Code est l'Ethique, et l'Ethique est le Code ».
- Pistes de réflexions : Charte éthique sur l'IA, Pluridisciplinarité, Formation, Interactivité, Gouvernance, etc.

Merci de votre attention



Jérôme Béranger (PhD) - Co-fondateur et CSO d'ADEL

jerome.beranger@ADEL-Health.com - 06 95 66 19 20

@jeromeberanger / @Label ADEL / www.adel-label.com