Il existe de nombreux types de moteurs, capables de traiter différentes formes de règles logiques pour déduire de nouveaux faits à partir de la base de connaissance.

On distingue souvent trois catégories, basées sur la manière dont les problèmes sont résolus :

Certains moteurs d'inférence[Lesquels ?] peuvent être partiellement pilotés ou contrôlés par des méta-règles qui modifient leur fonctionnement et leurs modalités de raisonnement.

Pour répondre à des objectifs spécifiques, ces moteurs d'inférence peuvent être doublés dans certains Systèmes Experts dits « inverses » ou « bimoteurs », ou encore être complétés par d'autres systèmes d'analyse pour constituer des Systèmes Experts dits « hybrides »[1].

Si les algorithmes de manipulation de faits et de règles sont nombreux et connus, la détermination de l'ensemble des faits et règles qui vont composer la base de connaissances est un problème délicat. Comment décrire le comportement d'un expert face à un problème particulier, et sa manière de le résoudre, là est la question. Car ce que l'on souhaite obtenir n'est ni plus ni moins que l'expérience, la connaissance pratique de l'expert, et non la théorie que l'on peut trouver dans les livres ni exclusivement les règles logiques d'inférence. Équivalents des méthodes d'analyse de l'informatique traditionnelle, des méthodes d'acquisition des connaissances sont développées.

Les systèmes d'apprentis

L'éthique de l'intelligence artificielle est le sous-domaine de l'éthique de l'informatique propre à l'intelligence artificielle (IA). Elle recouvre divers thèmes présentant des enjeux éthiques particuliers, dont les biais algorithmiques, l'équité, la prise de décision automatisée, la protection des données personnelles, l'imputabilité et la réglementation. L'éthique de l'IA recouvre également divers enjeux émergents ou potentiellement futurs tels que la sûreté et l'alignement de l'IA, la course aux armements, les systèmes d'armes létales autonomes, le chômage technologique, la désinformation, comment traiter certains systèmes d'IA s'ils ont un statut moral (bien-être et droits de l'IA), la superintelligence artificielle et les risques existentiels[1]. Certains domaines d'application de l'IA peuvent aussi avoir des implications éthiques particulièrement importantes, comme la santé, l'éducation ou le militaire.

L'émergence de l'intelligence artificielle (IA) soulève deux préoccupations majeures autour de la vie privée : des informations privées peuvent être compromises à travers une attaque contre un système d'IA par une tierce personne et un système d'IA lui-même peut être utilisé comme outil pour collecter des données privées sur des individus.

On trouve de nombreux exemples d'attaques contre un système d'IA par une tierce personne dans le domaine de l'apprentissage machine. En effet, les algorithmes d'apprentissage machine apprennent sur des données qui contiennent parfois des informations privées sur des individus, comme des informations médicales. L'entrainement de cet algorithme d'apprentissage machine produit ensuite un modèle qui peut être utilisé à travers des requêtes pour produire de nouvelles données. Il est donc nécessaire que les données d'entrainement ne puissent pas être reconstruites ou retrouvées à travers ces requêtes faites au modèle entrainé[2].

La recherche scientifique a permis de mettre en lumière plusieurs attaques permettant de reconstruire des données utilisées pour entrainer le modèle d'apprentissage machine à partir de requêtes au modèle[3]. Par exemple, il est possible de récupérer des clés de chiffrement privées en utilisant des outils de complétion de code basés sur l'apprentissage machine comme GitHub Copilot[4]. D'autres recherches ont démontré qu'il était possible de savoir si un individu fait partie d'une base de données d'entrainement sans y avoir directement accès[5],[6]. Cela peut porter atteinte à la vie privée de l'individu dans le cas où le simple fait de faire partie de la base de données est une information compromettante, par exemple lorsque la base de données regroupe des patients atteints d'une maladie grave[3].

De nombreuses technologies d'IA inquiètent les chercheurs quant à leur capacité à porter atteinte à la vie privée. Par exemple, les technologies de surveillance de masse comme la reconnaissance faciale peuvent représenter un danger pour la vie privée des individus[7].