

L'éthique en Intelligence Artificielle – Vincent Rialle

Résumé de mon allocution donnée le 18 mars 2025 lors du 6^{ème} Printemps éthique de l'UIAD à Grenoble, et reproduit dans les Actes du 6^{ème} Printemps éthique

Consacré à la découverte de la place de l'éthique en Intelligence artificielle (IA), cet exposé est structuré en quatre volets principaux.

- **Le premier volet** introduit la notion d'éthique, avec son sens, les valeurs qui la composent, les principes qui en découlent et la grande diversité de ses déclinaisons dans de nombreux champs d'application. Définie dans ses origines par Aristote, elle a été magistralement développée au XVII^e s. par Spinoza, suivie de nombreux autres penseurs jusqu'à nos jours. Compte tenu de la profusion de questions éthiques soulevées par notre civilisation, elle a été abondamment explicitée, popularisée et fait l'objet de saisissantes phrases pédagogiques présentant en quelques mots sa nature, ce qu'elle vise. Nous retiendrons ici celle de Paul Ricoeur : "Appelons « visée éthique » la visée de la « vie bonne » avec et pour autrui dans des institutions justes". L'éthique se situe donc au cœur de nos vies, individuelles, relationnelles et collective, jusqu'aux stades les plus complexes (institutions, organisation planétaire...). Prenons l'exemple de la bioéthique, composé des quatre valeurs fondamentales « bienfaisance, non-malfaisance, autonomie et justice ». Précisément déclinées dans le champ médical, elles sont aussi universelles : la non-malfaisance par exemple (abstinence de créer ou augmenter tout mal, souffrance, douleur) se retrouve dans toutes les autres formes ou champs d'application de l'éthique, mais aussi au cœur de nombreuses traditions depuis la nuit des temps. Le développement en France de la bioéthique en montre aussi sa fécondité « irradiant » en quelque sorte tous les domaines, jusqu'à celui du numérique.

- **Le deuxième volet** développe la vaste question « en quoi et pourquoi l'IA pose des problèmes éthiques ? » compte tenu de nombreux travers, scandales, dérives, attribués peu ou prou à l'IA elle-même – avec ses algorithmes omniprésents sur Internet, dans les réseaux sociaux, etc. – ainsi qu'au développement démesuré des « Big Tech », sociétés dotées de pouvoirs exorbitants allant jusqu'à celui d'ébranler fortement les démocraties. Avec l'explosion en 2022 du « phénomène chatGPT », la sidération provoquée par les extrêmes capacités langagières de l'IA est arrivée à son maximum suscitant de nombreuses réactions et initiatives, à visée typiquement éthique, du niveau local au niveau international. Localement par exemple, une équipe de chercheurs en IA de l'Université Grenoble Alpes, du MIAI¹ et du CNRS, a proposé au grand public (sur Youtube, à la Maison du Tourisme de

¹ Multidisciplinary Institute in Artificial Intelligence

Grenoble) une simulation de procès de l'IA selon un mode théâtralisé et humoristique, mais dont les contenus sont entièrement fondés sur des faits et chiffres réels et vérifiés, et intitulée *l'IA est-elle notre amie ? Le grand procès de l'IA (épisode 3)*. La chaire "Ethique & IA" du MIAI en ce qui concerne le champ académique, ainsi que le Lab d'éthique de l'UIAD, complètent ces exemples d'une intense activité éthique locale que l'on retrouve en d'innombrables villes, et même villages, en Europe et dans le monde.

- **Le troisième volet** retrace, en un panorama synthétique, des actions concrètes à visée éthique, aux niveaux national et international. Les Nations Unies, avec la phase d'analyse de la situation mondiale de l'IA lancée par l'ONU en 2019, ont publié en 2021 via l'UNESCO le rapport « Recommandations sur l'éthique de l'Intelligence Artificielle » signé par 193 pays, créé ensuite un « Observatoire mondial de l'éthique et de la gouvernance de l'IA » et réalisé le récent « Sommet pour l'Action sur l'IA » des 10 et 11 février 2025 à Paris, avec de nombreux chefs d'états présents ou représentés. Cet événement majeur – coprésidé par la France et l'Inde – a permis la création d'une fondation (*Current AI*) et la formulation d'un accord commun à une soixantaine de pays incluant la Chine (mais hors États-Unis et Royaume-Uni), plus l'Union européenne et la Commission de l'Union africaine, pour "une IA « ouverte », « inclusive » et « éthique »". L'Europe pour sa part connaît un véritable bon en avant depuis quelques années, avec notamment l'entrée en application en 2018 d'un Règlement général sur la protection des données (RGPD) puis, plus récemment celle du Règlement sur l'Intelligence artificielle (AI Act) visant à garantir, dans l'Union européenne, le respect des droits fondamentaux des citoyens et celui des « normes éthiques » par les systèmes d'IA. Ces exemples majeurs illustrent parfaitement le cheminement éthique résumé dans la phrase de Paul Ricœur.

- **Le quatrième volet** concerne les perspectives incertaines de ces engagements et le rôle moteur des citoyens du monde. Si le volet précédent évoque à n'en pas douter, une étape véritablement historique du cheminement éthique de l'IA, la réalité mondiale est celle d'un état d'esprit situé à rebours de l'évolution éthique. « La lutte entre nations pour la supériorité en matière d'IA causera probablement la troisième guerre mondiale » Elon Musk en 2017. Si nous n'en sommes pas tout-à-fait là, c'est bien par cette évolution éthique. Mais rien n'est assuré. Un esprit de féroce compétition prévaut toujours entre grands blocs nationaux pour une course à la suprématie de l'IA – pour la suprématie de l'armement et des moyens financiers qu'elle promet –, entre régimes dictatoriaux et démocraties, appétits démesurés des supers profits des Big Tech et, l'intérêt commun tel que défini notamment dans la « Déclaration sur l'avenir de l'internet » de 2022 en vue « *d'assurer que les valeurs que nous défendons hors ligne seront également protégées en ligne, de faire de l'internet un lieu sûr et un espace digne de confiance pour tous et de veiller à ce que l'internet serve notre liberté individuelle. Parce que l'avenir de l'internet est également celui de la démocratie, celui de l'humanité.* » (U. von der Leyen). Sachant que chacun des clics qu'individuellement

nous transmettons via nos écrans numériques sur certains « réseaux sociaux » contribue à une exorbitante puissance de déstabilisation (de certaines sociétés *Big Tech* largement à la botte de despotes dépourvus d'humanité), il s'agit aujourd'hui de travailler à bien penser, selon l'une des définitions essentielles de l'éthique, et agir en conséquence afin de rompre avec les captures d'attention systématiquement exercées par des algorithmes sur nos écrans numériques, laisser ainsi place aux relations créatrices de sens, autonomie et co-construction, et laisser ainsi ouvert l'avenir.
