

Résumé notes de cours

Fiches de révisions

En partie généré par IA à partir de la prise de note du cours

Attention à vérifier les informations

Une histoire de la philosophie

Intelligence(s) de l'Homme et de la Machine

Jean-Baptiste DUSSERT | EFREI | Année universitaire 2024/2025

1.Problématiques abordées

**Quelles sont les questions dont il faut avoir saisi le champ de la réponse dans son ensemble*

Dans quelle mesure l'IA impose-t-elle de repenser la nature de l'intelligence humaine dans son unicité et sa diversité ?

Crainte et résilience face au progrès technologique	<ul style="list-style-type: none">L'histoire montre une réticence face au progrès technologique (ex. Sainte Beuve, Heidegger).IA perçue avec méfiance, comparée aux bouleversements induits par les révolutions industrielles.Résonance des critiques de gauche (Marx) et de droite (Heidegger) sur les conséquences sociales et culturelles du progrès technologique.
Diversité de l'intelligence humaine	<ul style="list-style-type: none">Sujet traité par Howard Gardner avec ses 8 types d'intelligences (linguistique, logico-mathématique, intra-personnelle, etc.).IA possédant certaines formes d'intelligence (linguistique, logico-mathématique) mais manquant de conscience de soi, d'émotion réelle, etc.
Critiques de l'intelligence artificielle	<ul style="list-style-type: none">Intelligence de l'IA souvent simulée, rendant le concept d'intelligence distinct de la véritable compréhension (cf. Turing, Searle).La simulation de l'intelligence à travers l'accumulation de données (ex. hallucinations, IA limitée par le traitement syntaxique sans sémantique).
Impact sociétal et philosophique de l'IA	<ul style="list-style-type: none">La conception cartésienne de l'intelligence est remise en question (Descartes, Dennett).Tension entre la perception linguistique de l'intelligence et la réalité des capacités computationnelles.L'IA invite à reconsidérer le rôle des émotions et de la rationalité dans la prise de décision.
Nature, artifice et authenticité	<ul style="list-style-type: none">La distinction entre l'artificiel et le naturel selon Baudelaire, et son impact sur la perception de l'IA.L'IA comme une expression du technologiquement artificiel qui peut altérer les facultés humaines, posant des questions éthiques et philosophiques.
Symbolisme et perception du soi	<ul style="list-style-type: none">Le parallèle avec l'histoire des machines (ex. le Turc mécanique) souligne la tension entre performance apparente et habilité authentique.Rejet potentiel du progrès scientifique et retour à une vision où le langage est central mais non suffisant pour définir l'intelligence.
Redéfinition de l'humain à travers le prisme de l'IA	<ul style="list-style-type: none">L'idée que l'IA pourrait conduire à une redéfinition de l'intelligence humaine par sa capacité à simuler certains aspects cognitifs.Défis posés par l'IA par rapport à la hiérarchisation traditionnelle de l'intelligence (cf. anthropocentrisme).

Quelles sont les conséquences potentielles et déjà à l'œuvre de cette révolution, à la fois sur nos modes de vie et sur les valeurs fondamentales de la société, notamment le travail et la liberté ?

Résistance au progrès technique	<ul style="list-style-type: none">• Craintes historiques face à l'innovation (ex. Sainte-Beuve et la "littérature industrielle").• Résistance technologique de gauche : critiques de Marx, Fourier, Saint-Simon concernant l'impact du progrès sur le prolétariat.• Critique de droite : influence de Heidegger sur la technique affectant le rapport homme-monde, déshumanisation et aliénation culturelle.
Travail et économie	<ul style="list-style-type: none">• Transformation de la nature du travail, automatisation menaçant des emplois traditionnels.• Déclin de l'influence des ingénieurs et progression de la technologie perçue comme stagnante (Philippe Dessertine).• Exemples de résilience et d'innovation : robotisation, IA dans les services financiers et de santé.
Technologie et pouvoir	<ul style="list-style-type: none">• Biais décisionnels, asymétries de pouvoir dans l'usage de l'IA (Floridi).• Machines utilisées pour l'asservissement dans le contexte capitaliste.• Risque de contrôle de la société par des entités exploitant l'illusion de l'intelligence (John Searle's "Chambre chinoise").
Liberté et autonomie	<ul style="list-style-type: none">• Dépendance croissante aux technologies et réduction de l'autonomie humaine (critiques contemporaines du progrès).• Hans Jonas et le "Principe de responsabilité" : nécessité de questionner et d'encadrer les innovations pour éviter des conséquences négatives imprévues.• Considération des technologies comme moyen d'augmenter ou de restreindre la liberté humaine.
Identité et humanité	<ul style="list-style-type: none">• Problématique de l'identité humaine face à une intelligence artificielle compétitive (transhumanisme, mind uploading).• Prise de conscience des limites humaines et redéfinition de l'intelligence (Turing, Dennett, Damasio).• Humanoïdes et Uncanny Valley (Masahiro Mori): interroger les préjugés et perceptions humaines face aux machines humaines.
Société et valeurs	<ul style="list-style-type: none">• Impact culturel et social ambigu du progrès technique sur les valeurs traditionnelles et la société (Heidegger, Jonas).• Peur de l'artificialisation de l'humain dans un monde dominé par la technologie.• Inquiétude vis-à-vis de la déshumanisation et la réification de l'intelligence.

2. Concepts clefs

**Quels concepts doivent être maîtrisés pour pouvoir organiser la logique de sa réflexion*

Conscience

Heidegger et la technique	<ul style="list-style-type: none">• Heidegger critique la technique moderne qui éloignerait l'humain de son rapport authentique à l'être. Cette réflexion s'applique à l'IA qui simule l'intelligence sans conscience
Conception cartésienne de la conscience	<ul style="list-style-type: none">• Descartes lie conscience et pensée. L'IA, manquant de subjectivité, remet en question cette vision
Turing et illusion d'intelligence	<ul style="list-style-type: none">• L'IA peut simuler la conscience, mais selon Turing, cela reste une illusion car elle ne pense pas véritablement
Éthique du progrès technologique	<ul style="list-style-type: none">• Hans Jonas et le principe de responsabilité questionnent les conséquences de l'IA sur la conscience humaine
Dichotomie pensée/émotion	<ul style="list-style-type: none">• Dennett suggère que l'IA pourrait forcer une révision de la distinction traditionnelle entre raison et émotion

Éthique

Ethique et Technologie	<ul style="list-style-type: none">• Crainte des externalités négatives de l'IA, influence historique des critiques technologiques, gauche (Marx, Fourier, etc.) et droite (Heidegger)
Principe de Responsabilité (Hans Jonas)	<ul style="list-style-type: none">• Importance de questionner innovations, lien avec principe de précaution (Chirac, 2000s)
Critique de la Technologie	<ul style="list-style-type: none">• Progrès technique lié au capitalisme : exploitation humaine, déshumanisation.• Peur de l'aliénation et impact sur la conception de nous-mêmes, lié à l'illusion de maîtrise par l'IA

Précaution et responsabilité

Précaution technologique	Intellectuels expriment des craintes historiques face à l'innovation (référence à Sainte-Beuve et critiques historiques de gauche/droite)
Responsabilité philosophique et sociale	Importance de questionner les conséquences techniques (cf. Hans Jonas et le "Principe de responsabilité")
Heidegger et la critique technologique	La technologie perturbe le rapport naturel de l'homme au monde
Impact sur société et environnement	Depuis 2000, l'intégration du principe de précaution dans la politique française
Décroissance et scepticisme moderne	Méfiance envers le progrès, politique d'abstention

Intelligence

Origines du Concept	<ul style="list-style-type: none">Le terme "intelligence" vient du latin "intelligere" (inter + legere), impliquant la capacité de discerner et de faire des choix
Langue et Connotations	<ul style="list-style-type: none">En anglais, "intelligence" inclut l'idée d'accumulation de données, en plus du discernement
Critiques Historiques du Progrès	<ul style="list-style-type: none">Angst intellectuelle vis-à-vis des avancées techniques (Sainte-Beuve, Heidegger)
Types d'Intelligence	<ul style="list-style-type: none">Howard Gardner propose des intelligences multiples (linguistique, logico-mathématique, etc.), dont certaines sont simulées par l'IA
Illusion de l'IA	<ul style="list-style-type: none">L'IA donne une illusion d'intelligence par la manipulation de données, sans compréhension véritable
Test de Turing	<ul style="list-style-type: none">Questionne l'intelligence basée sur la communication humaine et la simulation
Menace Anthropocentrique	<ul style="list-style-type: none">Redéfinit comment nous percevons notre propre intelligence face à la machine

Machine

Anxiété envers la technologie	<ul style="list-style-type: none">• Sainte-Beuve (1839) et l'idée que la technologie annonce la fin d'une ère
Critique de la technique	<ul style="list-style-type: none">• Heidegger et son aversion pour le progrès technique qui altère le rapport naturel de l'homme au monde et critique liée au nazisme
Principe de responsabilité	<ul style="list-style-type: none">• Hans Jonas, principe précautionnaire en matière de technologie
Turing	<ul style="list-style-type: none">• Proposa que les machines pourraient simuler l'intelligence via un test
Chambre chinoise de Searle	<ul style="list-style-type: none">• Critique de la simulation d'intelligence
Heidegger et l'être	<ul style="list-style-type: none">• Différencie entre l'analyse linguistique et le calcul technique
Humanoïde et uncanny valley	<ul style="list-style-type: none">• Robots qui prennent forme humaine et leur rejet instinctif

Raison et émotion

Interactions Philosophiques	<ul style="list-style-type: none">• Damasio et Descartes explorent la tension entre rationalité et émotion. Pour Damasio, l'émotion est centrale dans la prise de décision rationnelle, remettant en question l'idée que la pensée rationnelle pure est idéale
Concept d'Expérience et de Perception	<ul style="list-style-type: none">• Dennett et la critique de la conscience cartésienne montrent comment la conscience et l'intellect ne se limitent pas à un processus centralisé, mais sont influencés par des émotions
Impact IA	<ul style="list-style-type: none">• La conception cartésienne traditionnelle est imitée dans l'IA, surtout dans la manière dont elle essaie de simuler des processus rationnels sans accorder autant d'importance aux émotions. Cela limite l'efficacité des IAs à n'égaler que la rationalité et négliger la dimension émotionnelle, essentielle pour une intelligence complète
Critiques de l'Approche Rationnelle Pure	<ul style="list-style-type: none">• Heidegger et d'autres proposent des visions où les techniques intellectuelles modernes, centrées sur des processus de calcul, manquent d'éléments essentiels comme l'expérience émotionnelle et la compréhension véritable.

Transhumanisme

Définition du transhumanisme	<ul style="list-style-type: none">Mouvement prônant l'usage des technologies pour augmenter les capacités humaines, notamment via IA, biotechnologies et nanotechnologies
Contexte historique	<ul style="list-style-type: none">Rejet et acceptation du progrès technologique. Influence de penseurs comme Heidegger et son disciple Hans Jonas sur la réflexion technologique
Philosophes clés	<ul style="list-style-type: none">Jacques Derrida remet en question l'écrit et la tradition occidentale. D'autres comme Descartes et La Mettrie empêchent de réduire l'humain à une machine
Questions éthiques	<ul style="list-style-type: none">Effets sur la nature humaine, responsabilité du progrès. Philosophie prudente face à l'innovation technologique
Applications IA dans le transhumanisme	<ul style="list-style-type: none">Neuralink, mind uploading

Travail

Historique et critique du travail	<ul style="list-style-type: none">• Évoque l'opposition des intellectuels face à l'industrialisation - crainte que la technologie ne change la nature du travail (Sainte-Beuve contre la "littérature industrielle")
Réactions politiques	<ul style="list-style-type: none">• Bifurcation idéologique (gauche vs droite) face à l'impact technologique sur le travail, avec Marx et Heidegger en figures critiques
Heidegger et la technique	<ul style="list-style-type: none">• Critique du progrès technique transformant négativement le rapport au monde, perçu dans des critiques de radio, industrie, et influence américaine
Principe de responsabilité et précaution	<ul style="list-style-type: none">• Hans Jonas et la réaction contre les conséquences potentielles des innovations technologiques, influençant des politiques comme le principe de précaution
Impact culturel	<ul style="list-style-type: none">• Réactions culturelles contre l'artificialisation du travail et de la vie humaine (ex. Descartes, La Mettrie)

3.Enjeux

**Quels sont les points sujets à débat qui pourraient faire l'objet d'une problématique*

Autonomie et dépassement de l'homme

L'autonomie et le dépassement de l'homme par l'IA suscitent des réflexions sur la transformation de notre compréhension de l'intelligence et du progrès technique.

Historiquement, les innovations technologiques, comme celles de la première révolution industrielle, ont généré des inquiétudes, en particulier concernant l'aliénation humaine et l'impact sur notre rapport au monde naturel, tel que critiqué par Heidegger.

L'IA, en tant que prolongation du progrès, pose la question de savoir si elle pourrait conduire à l'artificialisation de nos facultés humaines, altérant potentiellement notre condition naturelle et nous poussant à réévaluer ce qui fait de nous des êtres autonomes.

Nature de l'intelligence et diversité cognitive

La nature de l'intelligence et la diversité cognitive soulèvent des questions fondamentales dans l'ère de l'intelligence artificielle (IA). Historiquement, des craintes entourent l'innovation technologique, perçue comme menaçant notre compréhension unique de l'intelligence humaine.

La philosophie de **Heidegger** critique la technique, considérant que celle-ci altère notre lien naturel au monde. En parallèle, **Howard Gardner** propose des intelligences multiples, dont certaines déjà simulées par l'IA, comme les intelligences logico-mathématique et linguistique.

Cependant, la profondeur de la conscience humaine, de l'intelligence intra et interpersonnelle, reste un défi majeur non résolu pour l'IA.

Conscience et subjectivité artificielle

L'enjeu de la conscience et de la subjectivité artificielle soulève des questions profondes sur notre compréhension de l'intelligence et de l'existence même.

Descartes et **Heidegger** offrent des perspectives divergentes sur le sujet; **Descartes** voit l'intelligence comme le propre de l'humain pensant, tandis que **Heidegger** implique que notre rapport à l'être est fondamental pour notre identité.

Avec l'essor de l'intelligence artificielle, qui simule notre langage et apprentissage, nous devons examiner si cette simulation est juste un artifice vide ou si elle peut développer une conscience propre. Cette introspection pourrait transformer la manière dont nous percevons notre intelligence et notre identité.

Langage et compréhension réelle

Le langage joue un rôle central dans la question de la compréhension réelle, notamment face à l'IA. Historiquement, à travers des figures comme **Turing** et ses réflexions sur l'"**Imitation Game**", le langage est considéré comme une manifestation clé de l'intelligence.

Cependant, une IA, bien qu'elle puisse simuler une conversation humaine en accumulant et manipulant des données, demeure limitée. Elle opère principalement au niveau syntaxique sans véritable sémantique, créant ainsi une **illusion d'intelligence**.

Cette simulation pose des questions sur notre perception de la compréhension réelle et l'impact de l'IA sur notre définition de l'intelligence.

Rupture épistémologique et redéfinition du savoir

épistémologique : qui se rapporte à l'acte de connaissance scientifique. La tâche propre de l'épistémologie est donc herméneutique et historico- critique ; elle consiste à faire apparaître des organisations de concepts, qu'elles soient achevées ou imparfaites, des difficultés, ou obstacles, ou incohérences, des ouvertures, des points "sensibles"

Herméneutique : Théorie, science de l'interprétation des signes, de leur valeur symbolique. Appelons herméneutique l'ensemble des connaissances et des techniques qui permettent de faire parler les signes et de découvrir leur sens

La révolution de l'intelligence artificielle (IA) pose une rupture épistémologique significative, redéfinissant le savoir. Historiquement, les innovations technologiques ont suscité des craintes et résistances, notamment des intellectuels des deux bords politiques, préoccupés par les conséquences sociales et environnementales.

Avec l'IA, le savoir basé sur l'accumulation de données remplace partiellement l'intuition humaine. Ce changement provoque une tension : la connaissance devient quantitative, détachée de la subjectivité, redéfinissant notre rapport au monde et menaçant le paradigme traditionnel du savoir comme processus de compréhension humaine, un débat fondamental pour la philosophie moderne.

Rationalité, émotion et prise de décision

L'enjeu de la rationalité, de l'émotion et de la prise de décision dans le cadre de la philosophie de l'intelligence humaine et artificielle interroge la capacité des individus et des machines à prendre des décisions. Tandis que la rationalité est associée à la logique et à l'objectivité, les émotions, souvent perçues comme irrationnelles, influencent profondément les choix humains. Ce contraste soulève la question de savoir si les machines peuvent simuler cette complexité émotionnelle ou si la prise de décision purement rationnelle, comme dans l'intelligence artificielle, peut réellement remplacer celle humaine, souvent plus nuancée et contextuelle.

Responsabilité et éthique des technologies autonomes

L'enjeu de la responsabilité et de l'éthique des technologies autonomes concerne la question de savoir **qui est responsable des actions des machines autonomes**, comme les voitures sans conducteur ou les robots militaires. Si ces technologies prennent des décisions de manière autonome, cela soulève des **défis moraux et légaux** : qui assume la responsabilité en cas d'erreur ou de dommage ? Les machines peuvent-elles être éthiquement responsables ? Cela interroge aussi la manière dont les concepteurs et les sociétés encadrent l'utilisation de ces technologies pour garantir qu'elles respectent des principes éthiques tout en minimisant les risques.

Asymétries de pouvoir et biais algorithmiques

L'enjeu des asymétries de pouvoir et des biais algorithmiques soulève la question des inégalités renforcées par les technologies. Les algorithmes, souvent perçus comme neutres, peuvent reproduire ou accentuer des discriminations liées à des biais présents dans les données utilisées pour les entraîner. Ces biais peuvent affecter des domaines cruciaux, tels que la justice, l'emploi ou l'accès à des services essentiels, créant des discriminations invisibles mais réelles. Par ailleurs, la concentration du pouvoir entre les mains de quelques grandes entreprises technologiques soulève des préoccupations sur la gouvernance et le contrôle de ces outils, amplifiant les inégalités sociales et économiques.

Humanisation de la machine et déshumanisation de l'homme

L'enjeu de l'humanisation de la machine et de la déshumanisation de l'homme questionne l'impact des technologies sur l'identité et la société. L'intégration croissante de machines à l'apparence et au comportement humain, comme les robots sociaux ou les assistants virtuels, soulève des préoccupations sur l'authenticité des relations humaines. En parallèle, l'automatisation et l'usage excessif des technologies peuvent entraîner un éloignement de l'homme de ses propres capacités humaines, réduisant l'interaction sociale et la créativité individuelle. Cela pose la question de l'équilibre entre l'augmentation des capacités humaines par la technologie et la préservation de la dimension humaine dans nos vies quotidiennes.

Critères de l'intelligence et pertinence des tests

L'enjeu des critères de l'intelligence et de la pertinence des tests interroge la définition même de l'intelligence, qu'elle soit humaine ou artificielle. Les tests traditionnels, tels que le QI, mesurent des capacités spécifiques mais omettent souvent des dimensions essentielles de l'intelligence, comme la créativité, l'intelligence émotionnelle ou la capacité à résoudre des problèmes complexes. Lorsqu'il s'agit d'intelligence artificielle, la question se pose également de savoir si des tests conçus pour évaluer les machines peuvent vraiment capturer la richesse de l'intelligence humaine. Cela soulève des débats sur la validité des critères actuels et la nécessité de repenser les méthodes d'évaluation pour mieux refléter la diversité des intelligences.

4. Références

**Textes qui peuvent/doivent être cités pour appuyer sa réflexion*

ANDERS, Günther (philosophe allemand, 1902-1992)

L'Obsolescence de l'homme : Sur l'âme à l'époque de la deuxième révolution industrielle (1956), Paris, Ivrea, 2002.

L'IA constitue-t-elle le stade ultime de l'autosuffisance de la technique et d'une caducité de l'humanisme ?

Dans *L'Obsolescence de l'homme* (1956), Günther Anders aborde la question de l'impact de la révolution technologique, en particulier de l'industrialisation et des machines, sur l'humanité. Il soutient que l'humain, avec l'avènement de ces nouvelles technologies, devient obsolète, car il ne peut plus rivaliser avec l'efficacité des machines.

Le philosophe explore la manière dont les progrès technologiques déshumanisent l'individu et mettent en lumière l'incapacité de l'humanité à maîtriser les conséquences de sa propre création. La technique devient ainsi autosuffisante, au point où l'homme semble de plus en plus déconnecté de son propre destin et de sa capacité à comprendre et à contrôler les outils qu'il crée. Dans cette vision, l'IA ne constitue pas seulement un simple prolongement de la technique, mais un stade ultime où l'humain se trouve relégué à une position secondaire, voire obsolète.

En ce sens, Anders répond à la question posée en soulignant que la technique, et plus particulièrement l'IA, conduit à une "caducité" de l'humanisme, car elle rend l'homme inutile et incompetent face à l'autosuffisance de la machine. Il voit dans cette évolution un effacement de l'homme comme sujet autonome et une mise en cause profonde des valeurs humanistes. Ainsi, l'IA ne représente pas simplement une avancée technologique, mais une transformation radicale de la condition humaine.

BACHELARD, Gaston (philosophe français, 1884-1962)

L'avènement de l'IA doit-il être interprété en termes de rupture épistémologique, et si oui, comment ?

Dans *La Philosophie du non* (1947), Gaston Bachelard aborde la question de l'évolution des connaissances scientifiques et de la manière dont les paradigmes se transforment au fil du temps. Il introduit le concept de « rupture épistémologique », qui désigne le moment où une discipline ou une science abandonne un cadre de pensée ancien pour en adopter un nouveau, ce qui entraîne un changement radical dans la manière de comprendre le monde.

Bachelard répond à la question posée en interprétant l'avènement de l'IA comme une véritable rupture épistémologique. Selon lui, l'IA représente un tournant dans notre manière d'appréhender la connaissance et la rationalité humaines. Là où la science classique s'appuyait sur la pensée humaine comme source centrale de compréhension, l'IA, par son fonctionnement basé sur les algorithmes et les modèles computationnels, déplace cette source de rationalité vers les machines.

Cette rupture se manifeste par la remise en question des paradigmes traditionnels de la connaissance humaine. L'IA, en étant capable d'analyser des données à une échelle et avec une rapidité inaccessibles à l'esprit humain, modifie en profondeur notre conception de l'intelligence, du raisonnement et de la connaissance. Bachelard verrait dans l'émergence de l'IA un saut épistémologique majeur, redéfinissant les frontières de la science et de la cognition, et remettant en cause l'idée même d'une pensée exclusivement humaine.

En somme, pour Bachelard, l'avènement de l'IA est bien une rupture épistémologique qui transforme notre manière de concevoir la pensée, la connaissance et la rationalité, en plaçant la machine au cœur du processus cognitif.

DAMASIO, Antonio (neurologue luso-américain, né en 1944)

Si une adéquate compréhension mêle la rationalité et l'émotivité, une IA peut-elle prétendre rivaliser avec l'homme en ne reposant que sur des modèles purement computationnels ?

Dans son ouvrage *L'Erreur de Descartes* (1994), Antonio Damasio, neurologue luso-américain, explore la relation entre la rationalité et les émotions, et comment cette interaction est essentielle à la prise de décision humaine et à la conscience. Il soutient que la pensée humaine ne peut pas être réduite à de simples processus cognitifs ou computationnels, mais qu'elle nécessite une intégration complexe des émotions, des sensations corporelles et des processus rationnels.

Damasio répond à la question posée en affirmant qu'une IA, qui repose uniquement sur des modèles computationnels et purement rationnels, ne peut rivaliser avec l'humain, car elle néglige une dimension essentielle : l'émotion. Selon lui, la capacité humaine à raisonner et à prendre des décisions justes est intimement liée à l'expérience émotionnelle. Les émotions, loin d'être opposées à la raison, sont en réalité cruciales pour orienter et structurer nos pensées, nos choix et nos actions.

Ainsi, une IA basée uniquement sur des calculs et des algorithmes, sans intégrer cette dimension émotionnelle, serait incapable de reproduire la complexité et la profondeur de l'expérience humaine. En conclusion, Damasio souligne que, bien qu'une IA puisse simuler certains aspects du raisonnement humain, elle ne pourra jamais pleinement rivaliser avec l'homme tant qu'elle ne tiendra pas compte de l'importance des émotions et de leur interaction avec la rationalité.

DENNETT, Daniel C. (philosophe états-unien, né en 1942)

L'aboutissement à une IA douée de conscience ne suppose-t-il pas l'abandon de notre conception d'un « moi » central ?

Dans *La Conscience expliquée* (1991), Daniel Dennett explore la nature de la conscience et la manière dont elle émerge des processus cérébraux. Il défend une approche matérialiste et fonctionnaliste de l'esprit humain, affirmant que la conscience n'est pas un phénomène mystique ou non réductible, mais plutôt un produit des mécanismes cérébraux et des interactions complexes au sein du cerveau.

Dennett répond à la question posée en soutenant que l'aboutissement à une IA douée de conscience ne nécessite pas nécessairement l'abandon de notre conception du « moi » central, mais plutôt une révision de ce que nous entendons par ce « moi ». Selon lui, la conscience humaine est elle-même le résultat de processus décentralisés dans le cerveau, sans qu'il soit nécessaire d'avoir un "moi" unifié ou central. La conscience émerge d'un réseau complexe d'interactions entre des modules cérébraux spécialisés, et non d'un centre unique de contrôle.

Dans ce contexte, une IA dotée de conscience pourrait théoriquement fonctionner de manière similaire, sans avoir besoin d'un "moi" central au sens traditionnel du terme. L'idée d'un « moi » unifié et central, qui est souvent liée à des conceptions dualistes de l'esprit, pourrait ainsi être mise de côté dans une perspective fonctionnelle et matérialiste de la conscience. L'IA pourrait avoir une forme de conscience qui n'est pas fondée sur un centre unique, mais qui résulte d'une organisation complexe de processus.

En résumé, Dennett répond à la question en affirmant que l'émergence d'une IA consciente n'implique pas nécessairement l'abandon du concept de « moi », mais plutôt une redéfinition de ce concept, qui pourrait être compris comme un ensemble décentralisé de processus, et non comme une entité centralisée.

DESCARTES, René (philosophe français, 1596-1650)

Dans quelle mesure les notions d'animaux-machines et d'homme-machine (La Mettrie, 1748) continuent-elles d'influencer la manière dont nous concevons l'IA et nous mesurons à elle ?

Dans *Discours de la méthode* (1637), Descartes établit l'idée que les animaux sont des machines, dépourvues de pensée et d'âme, contrastant avec l'homme, doté de raison et d'âme. Cette distinction influence profondément la conception moderne de l'IA. Si Descartes voit l'homme comme supérieur grâce à sa pensée consciente, des penseurs comme La Mettrie, dans *L'Homme-machine* (1748), remettent en cause cette hiérarchie, assimilant l'homme à une machine. Cette vision continue d'influencer l'IA moderne, en opposant machines et consciences humaines, et en questionnant notre capacité à mesurer l'IA en fonction de la complexité humaine.

FLORIDI, Luciano (philosophe italien, né en 1964)

À quels biais décisionnels et asymétries de pouvoir devrions-nous être conscients de faire face dans l'usage de l'IA ?

Dans *La Quatrième Révolution* (2014), Luciano Floridi explore l'impact de l'IA et de l'informatique sur la société. Il met en lumière les biais décisionnels et les asymétries de pouvoir qui peuvent surgir de l'utilisation de l'IA. Floridi souligne que l'IA peut amplifier les biais humains présents dans les données utilisées pour entraîner les systèmes, ce qui peut conduire à des discriminations ou à des décisions injustes. De plus, il alerte sur la concentration du pouvoir entre les mains de ceux qui contrôlent les technologies de l'IA, ce qui pourrait exacerber les inégalités et nuire à la démocratie.

GARDNER, Howard (psychologue états-unien, né en 1943)

La dimension logico-mathématique des recherches actuelles en matière d'IA suffit-elle à rivaliser avec l'intelligence humaine ?

Dans *Les Esprits multiples* (1983), Howard Gardner explore l'idée que l'intelligence humaine est multidimensionnelle, comprenant plusieurs formes distinctes, telles que l'intelligence linguistique, spatiale, musicale et logique-mathématique. Il répond à la question en affirmant que la seule dimension logico-mathématique de l'IA ne suffit pas à rivaliser avec l'intelligence humaine, car l'humain possède une diversité de compétences et d'intelligences interconnectées. Selon Gardner, l'IA, même avancée dans le domaine logique, ne peut répliquer l'ensemble des capacités humaines, notamment celles liées à la créativité, à l'empathie et à la compréhension sociale.

JONAS, Hans (philosophe allemand, 1903-1993)

Le développement et l'usage de l'IA doivent-ils être freinés en regard de ses conséquences potentielles et le principe de précaution ne suggère-t-il pas intrinsèquement que nous sommes incapables d'anticiper de telles conséquences ?

Dans *Le Principe Responsabilité* (1979), Hans Jonas plaide pour une éthique de la responsabilité face aux technologies modernes, notamment l'IA. Il répond à la question en soutenant que le développement et l'usage de l'IA doivent effectivement être freinés en raison de leurs conséquences potentielles incalculables. Selon Jonas, le principe de précaution est fondamental, car il souligne que notre incapacité à anticiper pleinement les effets à long terme des nouvelles technologies, comme l'IA, nécessite une prudence accrue. L'incertitude quant aux risques futurs exige que nous agissions avec une responsabilité renforcée pour éviter des dangers irréversibles.

SEARLE, John (philosophe états-unien, 1932-2022)

La manipulation de symboles par une machine peut-elle véritablement être assimilée à une authentique intelligence et le langage naturel est-il la meilleure pierre de touche en la matière ?

Dans *La Lettre sur l'intelligence artificielle* (1980), John Searle explore la question de savoir si la manipulation de symboles par une machine peut être assimilée à une véritable intelligence. Il répond par la célèbre expérience de la "chambre chinoise", où une machine manipule des symboles sans compréhension véritable, montrant que la simple manipulation symbolique n'implique pas de conscience ou d'intelligence authentique. Searle soutient que l'intelligence réelle nécessite la compréhension et l'intentionnalité, et que le langage naturel est une pierre de touche cruciale, car il reflète la capacité humaine à saisir des significations au-delà de la simple syntaxe.

TURING, Alan (mathématicien britannique, 1912-1954)

Le test de Turing, héritier d'une longue tradition philosophico-mathématique, est-il pertinent pour juger de l'authenticité d'une intelligence ?

Dans *Computing Machinery and Intelligence* (1950), Alan Turing introduit le "test de Turing" pour évaluer si une machine peut imiter l'intelligence humaine. Il répond à la question en suggérant que l'authenticité de l'intelligence ne devrait pas être jugée par des critères philosophiques complexes, mais par la capacité de la machine à répondre de manière indiscernable d'un humain dans une conversation. Turing affirme que, si une machine réussit ce test, elle peut être considérée comme "intelligente", indépendamment de la manière dont elle fonctionne en interne, réorientant ainsi le débat de la philosophie vers des critères pratiques et observables.