



HAL
open science

L'injonction à la transparence : un levier réglementaire à double tranchant pour les organisations

Louis Vuarin, Véronique Steyer

► To cite this version:

Louis Vuarin, Véronique Steyer. L'injonction à la transparence : un levier réglementaire à double tranchant pour les organisations. Bulletin de l'Association Française pour l'Intelligence Artificielle, 2023, 120, pp.26-35. hal-04114847

HAL Id: hal-04114847

<https://hal.science/hal-04114847>

Submitted on 5 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



- cielle et action publique : construire la confiance, servir la performance, 31 mars 2022.
- [14] ETALAB. [Les algorithmes publics : enjeux et obligations.](#)
- [15] ETALAB. [Fiche pratique : l'obligation de mention explicite.](#)
- [16] E. Mouriessé. L'opacité des algorithmes et la transparence administrative. *Revue française de droit administratif*, page 45, 2019.
- [17] M. Saliou. [Les Français veulent des explications claires sur les modèles algorithmiques. Je peux parler à un humain ?](#), 22 février 2023.
- [18] T. Walsh. Turing's red flag – a proposal for a law to prevent artificial intelligence systems from being mistaken for humans. *Communications of the ACM*, 59(7) :34–37, 2016.
- [19] Wikipedia. [Locomotive Act.](#)

■ L'injonction à la transparence : un levier réglementaire à double tranchant pour les organisations

Par **Louis VUARIN**
I3 SES / Telecom Paris
Institut Polytechnique de Paris
louis.vuarin@telecom-paris.fr
louis.vuarin@polytechnique.edu

Véronique STEYER
I3 CRG / École Polytechnique
Institut Polytechnique de Paris
veronique.steyer@polytechnique.edu

Introduction

Réguler l'IA en exigeant davantage de transparence : un principe de plus en plus populaire, et qui semble se concrétiser progressivement au sein notamment de l'*AI Act* [1]. Pour autant, de telles injonctions à la transparence produisent-elles les effets attendus dans les organisations ? En particulier, la transparence est-elle vraiment un levier performant en matière de gestion des risques ? L'abondante littérature en sciences de gestion, et particulièrement en théorie des organisations appliquée à la gestion des risques, apporte un regard plutôt mitigé sur la question. Ces travaux alertent sur la possibilité de voir émerger une réglementation générant des dynamiques organisation-

nelles potentiellement contreproductives au regard des attentes affichées.

L'exigence de transparence : une panacée ?

De nombreux arguments plaident en faveur d'une exigence accrue de transparence pour les organisations qui conçoivent, commercialisent ou utilisent des IA. Pour lutter contre les biais et discriminations (au titre des origines socio-ethniques supposées, de critères géographiques, patrimoniaux, de genre, de préférences sexuelles, politiques, etc.) qui inquiètent la communauté académique et alimentent le débat sur l'éthique de l'IA [26, 30, 36, 51], la transparence offrirait une forme de garantie



de vérification et de correction des discriminations non souhaitées, qu'elles résultent des algorithmes ou des bases de données à partir desquelles ils sont entraînés. Sur le long terme, en accentuant les capacités de contrôle par les développeurs eux-mêmes ainsi que par des tiers externes [31, 34], la transparence renforcerait la confiance entre parties prenantes [17, 55]. En théorie, ce droit de regard, potentiellement renforcé par un protocole délibératif adéquat [10], provoquerait un cercle vertueux permettant de faire dialoguer et même converger utilisateurs et concepteurs, contribuant *in fine* à l'acceptabilité sociétale de cette technologie [13, 48, 49]. Fort de cet argumentaire, l'exigence de transparence s'est progressivement imposée comme un incontournable des chartes éthiques pour les organisations et méta-organisations qui se sont lancées dans ce type de démarche [20, 23, 25]. Dans leur méta-analyse de 84 chartes éthiques, Jobin *et al.* [23] identifiaient ainsi la transparence comme le principe le plus fréquemment mis en avant (73/84 documents), devant notamment la justice et l'équité (*justice & fairness*, 68/84), l'innocuité (*non-maleficence*, 60/84), la responsabilité (*responsibility*, 60/84) ou encore le respect de la vie privée (*privacy*, 47/84). Dans la même veine, la législation européenne reprend aussi ce principe, en l'intégrant graduellement comme une exigence réglementaire pour les organisations qui se lancent dans la conception ou la commercialisation d'IA. Ainsi, l'*AI Act*, dont une première mouture fut révélée par la Commission européenne en avril 2021[1], accorde une importance significative au principe de transparence [46]. Dans ses dispositions générales (Titre I), l'*AI Act* se fixe ainsi comme objet d'établir « des règles harmonisées en matière de transparence applicables aux systèmes d'IA destinés à interagir avec des personnes

physiques, aux systèmes de reconnaissance des émotions et de catégorisation biométrique, et aux systèmes d'IA utilisés pour générer ou manipuler des images ou des contenus audio ou vidéo » (article 1.d). L'article 13 précise les règles de « transparence et fourniture d'informations aux utilisateurs » des IA dites « à haut risque », et l'article 52 détaille les « obligations de transparence pour certains systèmes d'IA », par exemple pour informer les utilisateurs lorsqu'ils interagissent avec des IA (article 52.1), pour des IA utilisées dans la reconnaissance des émotions ou la catégorisation biométrique (article 52.2), ou encore pour encadrer l'utilisation d'IA à des fins de manipulation d'images, de sons et de vidéos (article 52.3)²⁵

Mais la transparence représente-t-elle vraiment une panacée face aux nombreux défis éthiques de l'IA ? Le succès du concept de transparence ne doit pas cacher certains raccourcis éthiques qui alimentent son argumentaire, et de nombreux auteurs mettent en garde envers les insuffisances théoriques de ce concept populaire [6, 7, 18, 20, 23, 25, 50]. Ces analyses développent deux lignes de critiques. En premier lieu, il s'agit de souligner le caractère imprécis du concept, qui tire son succès d'une métaphore facilement compréhensible à première vue, mais dont la définition est instable selon les acteurs qui la formulent [7, 25]. À cause de ce flou, les modalités d'application restent souvent à géométrie variable, peu contraignantes dans la pratique, avec dans certains cas le risque d'une stratégie d'affichage sans réel effet assimilable à une forme d'« *ethic washing* » [8, 32]. Par ailleurs, l'idéal de transparence se retrouve de facto limité par la technique [16, 26] : la plupart des algorithmes qui alimentent le boom actuel de l'IA sont structurellement opaques. Dès lors qu'ils ont été entraînés, ces algorithmes deviennent des « boîtes

25. Comme les « *deep fakes* », dont la création est de plus en plus facile et perfectionnée grâce à l'IA [5] mais dont l'encadrement par la loi pose de réelles questions juridiques et philosophiques [45].



noires » (*black boxes*) trop complexes pour être analysés dans le détail, sauf au prix d'efforts importants et avec une marge interprétative significative.

Mais au-delà des insuffisances du concept, il s'agit aussi d'explorer les effets secondaires de la transparence sur les organisations qui devront en appliquer les exigences. Car, si un idéal restant inefficace dans la pratique est dommageable pour ceux qui y investissent quelque forme d'espoir, il devient préoccupant pour la société dans son ensemble lorsqu'il s'avère produire des effets contre-intuitifs voire contre-productifs sur les acteurs et les organisations qu'il prétendait améliorer. Pour cela, il s'agit d'étudier l'impact organisationnel de l'injonction de transparence, et notamment d'anticiper les diverses dynamiques de jeux de pouvoir et de distorsions de l'attention qu'il peut produire. En effet, les sciences de gestion, notamment les auteurs qui se sont concentrés sur l'étude de la gestion des risques, ont montré que la transparence peut provoquer des conséquences inattendues sur le comportement des acteurs [2, 3, 4, 22, 38, 39, 40], obligeant à fortement relativiser les bénéfices éthiques attendus.

Mieux anticiper les effets de la transparence sur le comportement des acteurs dans les organisations

Dans cette perspective, nous proposons de mettre en valeur les deux faces de la médaille. Dans un premier temps, nous listons les dynamiques organisationnelles vertueuses qui sont suggérées par l'idéal de transparence appliqué aux organisations concevant ou commercialisant des IA. Puis, nous proposons de nous appuyer sur la littérature des sciences de gestion pour mieux caractériser les effets secondaires

qui pourraient résulter de l'application des injonctions à la transparence, telle que notamment formulée dans l'*AI Act*²⁶.

Côté pile : amorcer des dynamiques organisationnelles vertueuses

À première vue, l'*AI Act* semble élaboré selon une logique directement influencée par un objectif de diminution de l'ampleur et de la fréquence des conséquences défavorables de l'IA sur les personnes et les biens. Dans l'exposé des motifs de l'*AI Act*, la Commission européenne, tout en reconnaissant les nombreuses opportunités offertes par l'IA, met en garde contre le fait que « les éléments et techniques qui rendent possibles les bénéfices socio-économiques de l'IA peuvent aussi être à l'origine de nouveaux risques ou de conséquences négatives pour les personnes ou la société ». L'*AI Act* fait suite à de nombreux travaux préliminaires à l'initiative de différents organes nationaux et européens, qui ont souligné la nécessité d'une approche équilibrée qui encouragerait le développement d'un écosystème économique et industriel favorable à l'IA sans ignorer les conséquences sur les européens. Dans cette optique, la mouture du 21 avril 2021 prétend, dans son exposé des motifs, répondre « à des demandes explicites du Parlement européen et du Conseil européen, qui ont lancé plusieurs appels en faveur de l'adoption de mesures législatives visant à assurer le bon fonctionnement du marché intérieur des systèmes d'intelligence artificielle (ci-après les « systèmes d'IA ») en mettant en balance les bénéfices et les risques de l'IA à l'échelle de l'Union ».

Ni une jungle anarchique, ni un écosystème innovant étouffé par une législation inhibante : le législateur est en quête d'une troisième voix,

26. Certaines professions en pointe sur les conséquences de l'IA sur leur pratique du quotidien, comme l'audit et la médecine [29], mènent déjà un tel travail de prospective : notre lecture s'inscrit dans ce souci d'anticipation des risques liés à l'IA en l'élargissant aux organisations en général, en s'appuyant sur les travaux en théorie des organisations questionnant la gestion des risques et les effets de la transparence.



et d'outils concrets pour la mettre en œuvre. Dans cette quête d'équilibre réglementaire, la transparence est apparue comme le principal levier normatif applicable aux organisations développant et commercialisant des IA pour matérialiser cette obligation de rendre des comptes aux nombreuses parties prenantes affectées par leurs technologies. Au fil de ses 85 articles ainsi que de ses travaux préparatoires et préambules, l'*AI Act* établit ainsi un lien clair entre exigences de transparence et gestion des risques. Les obligations de transparence sont alors modulées en fonction du niveau de risque associé au type d'IA et à son usage. Schématiquement, l'argument central derrière cette gradation des exigences en fonction des types d'IA et de leur niveau de risque repose sur l'idée que les exigences de transparence agiraient comme des garde-fous pour les organisations qui développent et commercialisent des IA : en rendant les IA plus transparentes, elles deviendraient plus explicables, ce qui faciliterait le contrôle des acteurs. Implicitement, cette démarche repose sur l'espoir de produire trois dynamiques vertueuses. Premièrement, le contrôle rendu possible par les exigences de transparence devrait déjà produire une partie de ses fruits en obligeant les organisations et leurs acteurs à améliorer leurs processus de conception et de mise en usage en amont pour être irréprochables lorsque surviendrait l'audit. On pourrait appeler ce processus une amélioration par anticipation de la sanction. Deuxièmement, l'idéal de transparence suggère aussi que la transparence inviterait de manière corollaire l'organisation à adopter le point de vue de celles et ceux qui pourraient exiger un contrôle, et donc à développer son attention sur certains effets indésirables générés par les IA qu'elle conçoit ou qu'elle vend. On pourra ici parler d'amélioration par réflexivité organisationnelle. Enfin, en tout état de cause, le contrôle facilité par les obligations de transparence accélérerait la mise

en place de plans de remédiation dans le cas où certains problèmes seraient observés : il s'agirait d'une amélioration par gain de réactivité des actions correctrices.

Côté face : museler le doute, le tacite et l'informel, si cruciaux en matière de gestion des risques

Les efforts consentis par les organisations pour gagner en transparence se justifient donc par le gain apporté par une telle gouvernance au regard de la capacité de ces organisations à identifier et mitiger les risques en amont des contrôles (amélioration par anticipation, par réflexivité organisationnelle) et en aval (amélioration par gain de réactivité des actions correctrices). Mais dans la pratique, la littérature en gestion et de nombreuses études empiriques montrent que la transparence ne se traduit pas nécessairement par une meilleure gestion des risques. Parmi les principaux arguments invitant à la méfiance envers une application trop dogmatique de la transparence, on peut ainsi noter des effets d'occultation volontaire via la transparence ; l'effet censure de la transparence sur les savoirs tacites et informels ; et en corollaire de cette censure, l'effet rédhibitoire de la transparence sur l'expression du doute.

L'effet d'occultation par la transparence résulte d'une tentative de saturation de l'attention en inondant les observateurs d'information. Stohl *et al.* [42] distinguent entre l'opacité involontaire – dans laquelle « la visibilité produit de si grandes quantités d'informations que des éléments d'information importants sont cachés par inadvertance dans les débris des informations rendues visibles » – et l'opacité stratégique – dans laquelle les acteurs « liés par des règles de transparence » rendent volontairement tellement d'informations « visibles que les éléments d'information sans importance prendront tellement de temps et d'efforts à passer au crible que les récepteurs seront distraits de



l'information centrale que l'acteur souhaite dissimuler » (pp. 133-134).

L'effet de censure vient d'une forme d'opprobre des savoirs tacites et informels véhiculé par l'idéal de transparence dont les conséquences peuvent être plus insidieuses, et donc potentiellement plus difficiles à remédier.

L'idéal de transparence véhicule en effet l'idée que le management des risques est une affaire entièrement explicitable et objectivable. Un tel arc argumentatif se résume au fond à la sentence de Boileau : « Ce que l'on conçoit bien s'énonce clairement, Et les mots pour le dire arrivent aisément. » Dans cette logique, qualité, sécurité, éthique, répondent d'une même problématique : la nécessité d'exprimer, qui résulte forcément en une capacité à objectiver formellement le problème et ses solutions, et inversement à se méfier de toutes autres formes de gestion se fondant sur du ressenti ou de l'implicite. Or, la littérature en sciences de gestion depuis trois décennies au moins est particulièrement claire sur la nécessité de prendre en compte l'informel et le tacite comme des facteurs clé de la résilience des organisations [14, 21, 28, 41, 47]. De nombreux auteurs spécialisés sur le management des HRO (*High Reliability Organizations*²⁷) ont ainsi montré comment des outils de gestion supposés identifier et objectiver les risques provoquent parfois une myopie organisationnelle sur les vraies priorités, en muselant certaines expertises et connaissances des risques fondés sur le ressenti et les savoirs informels [27, 52, 53, 54].

De peur de ne pouvoir expliquer clairement leurs raisonnements, techniciens et managers pourraient avoir tendance à taire ce qui est de l'ordre du sensible (par exemple : « j'ai la drôle d'impression que ce résultat est faux, même si je ne sais pas dire clairement pourquoi ») L'im-

plicite hyper-rationaliste de l'idéal de transparence peut alors être contre-productif en matière de gestion des risques : sous son emprise s'opère un renversement des valeurs dans lequel le tacite est perçu comme la source du risque alors qu'il s'avère parfois pourtant crucial dans la gestion des risques : que ce soit pour permettre une identification précoces des anomalies et vulnérabilités ou pour y remédier [11, 47, 54]. Ainsi, sous le sceau de l'injonction à la transparence, les ajustements locaux, officieux, informels, sont considérés comme déviant et inacceptables : or, ils sont non seulement monnaie courante, mais même parfois franchement nécessaires à la bonne résilience des organisations [21, 47, 54]. Le bricolage et l'improvisation organisationnelle, même inexpliqués, contribuent à la bonne marche de l'organisation et au développement d'une certaine expertise [12]. En complément des plans « prescrits » de gestion des risques, la littérature souligne l'importance de respecter ce qui pourrait apparaître comme des arrangements locaux, des routines et rites, des mythes professionnels, voire des « bidouilles », du moment qu'ils participent effectivement d'une meilleure gestion de l'activité « réelle » [14, 21], sans pour autant qu'elle passe par une explicitation *in extenso* de ses tenants et aboutissants.

En mésestimant les savoirs tacites de la gestion des risques, l'injonction à la transparence pourrait alors détricoter une certaine expertise informelle en la matière, et aggraver la capacité des organisations à y faire face de manière adéquate. Mais il y a pire : corollairement au discrédit sur les savoirs tacites et informels, l'idéal de transparence peut tendre à assécher la pratique du doute. Or, être capable d'exprimer son doute, par exemple une crainte pour la sécurité du système fondée sur l'expé-

27. Les Organisations Hautement Fiables sont des organisations qui doivent parvenir à fonctionner dans des univers particulièrement complexes et incertains tout en réussissant à maintenir un haut niveau de fiabilité, comme les centrales nucléaires, les porte-avions, le contrôle aérien ou encore les urgences hospitalières.



rience et le ressenti, est une dimension critique de toute organisation résiliente [54]. Organiser cette reconnaissance et cette valorisation du doute nécessite non seulement d'entraîner les acteurs à identifier la sensation du doute et à le formuler à partir de leurs expériences et savoirs bien souvent tacites, mais aussi de disposer d'une culture organisationnelle et de circuits décisionnels aptes à traiter et à réagir au doute. C'est un enjeu du management de l'attention par définition bousculé, sinon écrasé, par l'idéal de transparence lorsqu'il est nourri par un présupposé hyper-rationaliste de la gestion des risques.

Pourquoi ces problématiques organisationnelles sont importantes en matière de management de l'IA ?

En matière d'intelligence artificielle, l'expertise informelle et l'ajustement progressif sont des dimensions clés d'amélioration et de robustesse de la conception [44].

Dans cette perspective, il convient de souligner l'importance de la dimension processuelle du développement d'IA : même lorsqu'elle suit des cycles de conception rapides, l'IA n'apparaît pas *ab nihilo*, et le résultat final est le cumul d'un certain nombre de décisions et de modifications dont il est impératif de garder la mémoire. En particulier, la phase d'apprentissage des algorithmes de machine learning est le fruit de successions d'ajustements et d'améliorations, guidés par un ensemble de métriques, mais aussi par de l'astuce, de l'expérience, et des jeux d'essais-erreurs qui oscillent entre ce que l'on appelle couramment les « ficelles du métier » [43] et du « bricolage » raisonné [9, 15, 37]. Les développeurs jouent sur les paramètres, les algorithmes, les bases d'apprentissage, avec une pratique relevant aussi bien des sciences computationnelles que d'un certain art.

À ce titre, ce qui compte n'est pas seule-

ment de retracer les ajustements que les *datascientists* ont pu réaliser, mais aussi de comprendre les ajustements qu'ils n'ont pas réalisés, et de comprendre pourquoi. Glaser *et al.* [19] décrivent ainsi l'IA comme un « assemblage », combinant savoirs explicites et implicites, techniques, mais aussi relations informelles, habitudes et routines, préférences organisationnelles, modalités du processus décisionnel, etc. Une fois conçue, la phase d'adoption oblige de nouveau à modeler l'IA en fonction d'un ensemble de processus formels et informels, techniques et sociaux, propres à l'organisation adoptante. Murray *et al.* [33] décrivent ainsi une « agence siamoise » (« *conjoined agency* ») entre homme et machine, et Kellogg *et al.* [24] et Neirotti *et al.* [35] mettent en lumière les nouveaux métiers liés à l'émergence des algorithmes dans les organisations dont les missions exigent précisément d'ajouter du liant et un certain sens de l'interprétation fonctionnelle pour que les outils algorithmiques s'insèrent plus facilement dans le tissu organisationnel. Cette lecture sociotechnique et processuelle de l'IA, et en particulier le rôle de l'expérience et des ajustements durant le processus de conception, sont des dimensions qui se retrouvent non seulement occultées par l'injonction à la transparence, mais qui pourraient même à terme subir une forme de censure organisationnelle si elles ne coïncident pas avec les modalités du contrôle envisagé.

Dans cette perspective organisationnelle, deux dimensions manquantes au principe de transparence semblent ainsi nécessaires d'être renforcées dans la loi. La première porte donc sur la reconnaissance de l'informel et du tacite, incluant l'expérience et l'expertise, et l'importance du doute et des essais-erreurs dans la gestion des risques. Si la métaphore de la « transparence », qui met l'accent sur l'objet technique, et non sur les qualifications et l'intelligence des hommes et des organisations,



est insuffisante, d'autres concepts pourraient lui être favorablement substitués, comme celui de sincérité et d'auditabilité. La seconde dimension porte sur l'effet cristallisateur du principe de transparence, qui a tendance à « fixer » les limites spatiotemporelles du contrôle [6] : l'organisation est obligée de produire des rapports de l'état du système à un instant t . En orientant l'attention de l'organisation sur le timing du contrôle, la réglementation peut passer à côté d'un renforcement des exigences sur le processus de conception, avec ses allers-retours, ses choix, ses hésitations, ses hypothèses, etc. La mémoire de ce processus est pourtant incontournable pour une bonne gestion des risques, mais elle devient relativement incompatible avec l'eschatologie imposée en substrat par le principe de transparence. *In fine*, c'est alors tout un ressort de la résilience des organisations et du professionnalisme de ses acteurs qui pourraient s'en trouver durablement affectés.

Références

- [1] Proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (dit « AI Act »), COM/2021/206 final, 21 avril 2021. [Voir la version actuelle de l'AI Act](#) : Orientation générale du Conseil de l'Union européenne modifiant la proposition d'AI Act, document ST 15698 2022 INIT, 6 décembre 2022.
- [2] M. Al Balushi. How internal transparency impacts organizational resilience. *International Journal of Quality & Reliability Management*, 38(5) :1246–1263, 2021.
- [3] O. B. Albu and M. Flyverbom. Organizational transparency : Conceptualizations, conditions, and consequences. *Business & Society*, 58 :268–297, 2019.
- [4] O. B. Albu and L. Ringel. The perils of organizational transparency : Consistency, surveillance, and authority negotiations. In *Toward Permeable Boundaries of Organizations ?*, pages 227–256. Emerald Publishing Limited, 2018.
- [5] A. M. Almars. Deepfakes detection techniques using deep learning : a survey. *Journal of Computer and Communications*, 9(5) :20–35, 2019.
- [6] M. Ananny and K. Crawford. Seeing without knowing : Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3) :973–989, 2018.
- [7] G. Andrada, R. W. Clowes, and P. R. Smart. Varieties of transparency : Exploring agency within ai systems. *AI & Society*, pages 11–11, 2022.
- [8] E. Bietti. From ethics washing to ethics bashing : a view on tech ethics from within moral philosophy. In *Conference on fairness, accountability, and transparency*, pages 210–219, 2020.
- [9] E. Boxenbaum and L. Rouleau. New knowledge products as bricolage : Metaphors and scripts in organizational theory. *Academy of Management Review*, 36(2) :272–296, 2011.
- [10] A. Buhmann and C. Fieseler. Towards a deliberative framework for responsible innovation in artificial intelligence. *Technology in Society*, 64 :101475, 2011.
- [11] D. Chernov, D. Sornette, G. Sansavini, and A. Ayoub. Examples of failures in intra-organizational risk transmission in past disasters. In *Don't Tell the Boss! How Poor Communication on Risks within Organizations Causes Major Catastrophes*, pages 13–332. Springer International Publishing, 2023.



- [12] F. Chédotel. L'improvisation organisationnelle. *Revue française de gestion*, 1 :123–140, 2005.
- [13] K. de Fine Licht and J. de Fine Licht. Artificial intelligence, transparency, and public decision-making : Why explanations are key when trying to produce perceived legitimacy. *AI & Society*, 35 :917–926, 2020.
- [14] M. Detchessahar, S. Gentil, A. Grevin, and B. Journé. Entre cacophonie et silence organisationnel, concevoir le dialogue sur le travail. le cas de projets de maintenance dans une industrie à risque. *Annales des Mines – Gérer et comprendre*, 4 :33–45, 2017.
- [15] R. Duymedjian and C. C. Rüling. Towards a foundation of bricolage in organization and management theory. *Organization Studies*, 31(2) :133–151, 2017.
- [16] L. Edwards and M. Veale. Slave to the algorithm? why a 'right to an explanation' is probably not the remedy you are looking for. *16 Duke Law & Technology Review*, 18 :2972855, 2017.
- [17] N. Emaminejad and R. Akhavian. Trustworthy AI and robotics : Implications for the AEC industry. *Automation in Construction*, 139 :104298, 2022.
- [18] H. Felzmann, E. F. Villaronga, C. Lutz, and A. Tamò-Larrieux. Transparency you can trust : Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6(1) :2053951719860542, 2019.
- [19] V. L. Glaser, N. Pollock, and L. D'Adlerio. The biography of an algorithm : Performing algorithmic technologies in organizations. *Organization Theory*, 2(2) :26317877211004609, 2021.
- [20] T. Hagendorff. The ethics of AI ethics : An evaluation of guidelines. *Minds and Machines*, 30(1) :99–120, 2020.
- [21] E. Hollnagel, B. Journé, and H. Laroche. La fiabilité et la résilience comme dimensions de la performance organisationnelle. *M@n@gement*, 12(4) :224–229, 2009.
- [22] M. Jendly. Performance, transparence et accountability : une équation (dé) responsabilisante des professionnels exerçant en prison? *Déviance et société*, 36(3) :243–262, 2012.
- [23] A. Jobin, M. Ienca, and E. Vayena. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9) :389–399, 2019.
- [24] K. C. Kellogg, M. A. Valentine, and A. Christin. Algorithms at work : The new contested terrain of control. *Academy of Management Annals*, 14(1) :366–410, 2020.
- [25] S. Larsson and F. Heintz. Transparency in artificial intelligence. *Internet Policy Review*, 9(2) :2020.2.1469, 2020.
- [26] K. Martin. Ethical implications and accountability of algorithms. *Journal of Business Ethics*, 160 :835–850, 2019.
- [27] J. Mayer. *De l'attention au risque : une perspective attentionnelle de la construction sociale du risque par les organisations*. PhD thesis, Université Paris-Dauphine – Paris Sciences et Lettres, 2017.
- [28] J. C. Mayer. Influencer l'attention des décideurs-les pratiques d'« issue-selling » des risk managers. *Revue française de gestion*, 42(255).
- [29] S. McLennan, A. Fiske, D. Tigard, R. Müller, S. Haddadin, and A. Buyx. Embedded ethics : a proposal for integrating ethics into the development of medical AI. *BMC Medical Ethics*, 23(1).



- [30] B. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi. The ethics of algorithms : Mapping the debate. *Big Data & Society*, 3(2) :2053951716679679, 2016.
- [31] M. Mora-Cantalops, S. Sánchez-Alonso, E. García-Barriocanal, and M. A. Sicilia. Traceability for trustworthy ai : A review of models and tools. *Big Data and Cognitive Computing*, 5(2) :20, 2021.
- [32] J. Morley, A. Elhalal, F. Garcia, L. Kinsey, J. Mökander, and L. Floridi. Ethics as a service : a pragmatic operationalisation of AI ethics. *Minds and Machines*, 31(2) :239–256, 2021.
- [33] A. Murray, J. E. N. Rhymer, and D. G. Sirmon. Humans and technology : Forms of conjoined agency in organizations. *Academy of Management Review*, 46(3) :552–571, 2021.
- [34] J. Mökander and L. Floridi. Operationalising AI governance through ethics-based auditing : an industry case study. *AI and Ethics*, s43681-022-00171-7, 2022.
- [35] P. Neirotti, D. Pesce, and D. Battaglia. Algorithms for operational decision-making : An absorptive capacity perspective on the process of converting data into relevant knowledge. *Technological Forecasting and Social Change*, 173 :121088, 2021.
- [36] E. Ntoutsis, P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdil, M. E. Vidal, S. Ruggieri, F. Turini, S. Papadopoulos, E. Krasanakis, I. Kompatsiaris, K. Kinder-Kurlanda, C. Wagner, F. Karimi, M. Fernandez, H. Alani, B. Berendt, T. Kruegel, C. Heinze, K. Broelemann, G. Kasneci, T. Tiropanis, and S. Staab. Bias in data-driven artificial intelligence systems – an introductory survey. *Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery*, 10(3) :e1356, 2020.
- [37] N. Phillips and P. Tracey. Opportunity recognition, entrepreneurial capabilities and bricolage : connecting institutional theory and entrepreneurship in strategic organization. *Strategic Organization*, 5(3) :313–320, 2007.
- [38] L. Ringel. Boundaries of visibility in the age of transparency : An integrative conceptualization. In *Toward Permeable Boundaries of Organizations ?*, pages 55–79. Emerald Publishing Limited, 2018.
- [39] L. Ringel. Unpacking the transparency-secrecy nexus : Frontstage and backstage behaviour in a political party. *Organization Studies*, 40(5) :705–723, 2019.
- [40] J. Spindler. La transparence de la gestion publique : de la recherche d'un plus grand approfondissement à un risque d'opacité. *Gestion Finances Publiques*, 6(6) :68–77, 2020.
- [41] A. Stimec and B. Journée. Faire face à la complexité par le dialogue technique et la négociation-le cas de la sûreté dans l'industrie nucléaire. *Revue française de gestion*, 47(297) :123–140, 2021.
- [42] C. Stohl, M. Stohl, and P. M. Leonardi. Managing opacity : Information visibility and the parado of transparency in the digital age. *International Journal of Communication*, 10 :123–137, 2016.
- [43] M. Stroobants. Dénouer les ficelles du métier pour connecter les savoirs formels et informels. *Techniques & Culture*, 51 :164–179, 2009.
- [44] L. A. Suchman and R. H. Trigg. Artificial intelligence as craftwork. In S. Chaiklin and J. Lave, editors, *Understanding practice : Perspectives on activity and context*, pages 144–178. Cambridge University Press, 1993.



- [45] B. Van der Sloot and Y. Wagenveld. Deepfakes : regulatory challenges for the synthetic society. *Computer Law & Security Review*, 46 :[105716](#), 2022.
- [46] M. Veale and F. Zuiderveen Borgesius. Demystifying the draft EU Artificial Intelligence act (july 31, 2021). *Computer Law Review International*, 22(4) :97–112, 2021.
- [47] T. J. Vogus, N. B. Rothman, K. M. Sutcliffe, and K. E. Weick. The affective foundations of high-reliability organizing. *Journal of Organizational Behavior*, 35(4) :592–596, 2014.
- [48] E. S. Vorm and D. J. Combs. Integrating transparency, trust, and acceptance : The intelligent systems technology acceptance model. *International Journal of Human-Computer Interaction*, 38(18–20) :1828–1845, 2022.
- [49] J. Wanner, L. V. Herm, K. Heinrich, and C. Janiesch. The effect of transparency and trust on intelligent system acceptance : Evidence from a user-based study. *Electronic Markets*, 32 :2079–2102, 2022.
- [50] R. Warner and R. H. Sloan. Making artificial intelligence transparent : Fairness and the problem of proxy variables. *Criminal Justice Ethics*, 40(1) :23–39, 2021.
- [51] A. L. Washington. How to argue with an algorithm : Lessons from the COMPAS-ProPublica debate. *Colorado Technology Law Journal*, 17(1) :[3357874](#), 2019.
- [52] K. E. Weick and K. M. Sutcliffe. *Managing the unexpected*. San Francisco : Jossey-Bass, 2001.
- [53] K. E. Weick and K. M. Sutcliffe. Mindfulness and the quality of organizational attention. *Organization Science*, 17(4) :514–524, 2006.
- [54] K. E. Weick and K. M. Sutcliffe. *Managing the unexpected : Resilient performance in an age of uncertainty*. John Wiley & Sons, 2007.
- [55] J. Zerilli, U. Bhatt, and A. Weller. How transparency modulates trust in artificial intelligence. *Patterns*, 3 :[100455](#), 2022.