

Compétence transversale – L1

Grands défis sociétaux : Intelligence Artificielle

Pr. Lucile Sassatelli

Professeure des Universités en Informatique, UniCA

Directrice scientifique de EFELIA Côte d'Azur

Image by Alan Warburton / © BBC / Better Images of AI / Nature / CC-BY 4.0

Plan du module

Chapitre	Titre	Contenu	Date d'ouverture	Date QCM
1	Qu'est-ce que l'IA ?	• Domaine et familles de méthodes • Applications actuelles • Définition d'un modèle d'IA, choix humains dans la conception • Mythe de l'AGI et confrontation à la réalité des neurosciences		
2	Qu'est-ce que les grands modèles de langue (LLM) ?	• Principe des modèles génératifs de texte • Choix simplificateurs et conséquences : faillibilité, biais		• QCM 1 noté 3-7/11
3	Quels impacts sur la société ? Quels impacts sur moi ?	• Exploitation humaine et emplois • Exploitation environnementale • Apprentissage à l'université		
4	Quels impacts sur ma discipline ?	• Les intérêts des systèmes d'IA • Les enjeux des systèmes d'IA pour son domaine		• QCM 2 note 8-12/12



Objectif

- Ce chapitre expose des exemples de **travaux scientifiques en lien avec l'IA pour chacun des 7 portails disciplinaires de licence à UniCA.**
- Ils présentent soit des avancées, soit des problèmes de déployer des SIA pour certains objectifs.
- Ceci pour vous donner des exemples pour amorcer votre compréhension du lien entre IA et votre discipline.



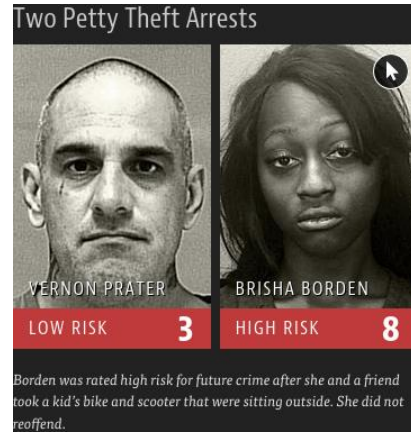
Plan

- 
1. Droit et science politique
 2. Economie-gestion
 3. Langues Littérature Arts et Création
 4. Médecine, psychologie, sciences de la vie
 5. Sciences humaines et sociales
 6. Sciences et Technologies
 7. STAPS



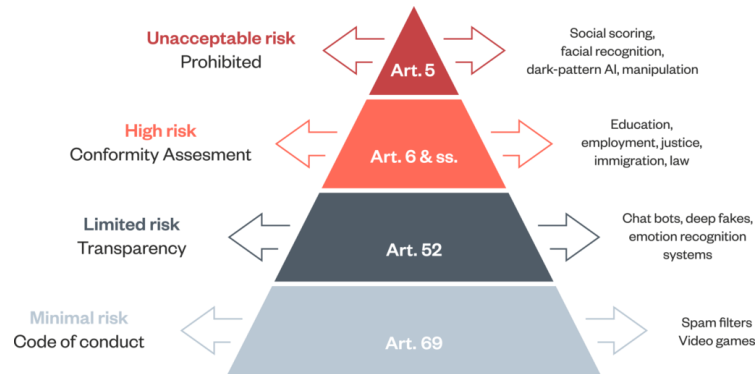
Droit et science politique

- Justice prédictive



[Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks](#)

- Règlement européen sur l'IA



<https://www.adalovelaceinstitute.org/resource/eu-ai-act-explainer/>
<https://artificialintelligenceact.eu/fr/>

- Justice

- J. Angwin et al., "[Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks.](#)", Propublica, 2016.

- Guerre

- Amnesty International, "[Israel and Occupied Palestinian Territories: Automated Apartheid: How facial recognition fragments, segregates and controls Palestinians in the OPT,](#)" May 2023.
- A. Iraqi, "['Lavender': The AI machine directing Israel's bombing spree in Gaza,](#)" +972 Magazine, Avril 2024.

- Démocratie

- OCDE, "[AI chatbots aren't reliable for voting information, government officials warn - OECD.AI.](#)" Nov. 2024.
- G. Simmons, "[Moral Mimicry: Large Language Models Produce Moral Rationalizations Tailored to Political Identity,](#)" ACL, 2023.

- Droit

- R. Cahill, "[OpenAI Defamation Lawsuit: The first of its kind,](#)" Syracuse Law Review, 2023.

- Outil

- V. Magesh et al., "[Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools](#)", NeurIPS, 2024.

Economie-gestion

- Gains de **productivité** initialement annoncés et prévisions de 27 % d'économies sur les coûts de main-d'œuvre. Ces prévisions sont révisées :
 - 4,6 % des tâches affectées par l'IA, augmentation du facteur de **productivité de 0,66 % sur 10 ans** (0,06 %/an)
 - augmentation de la croissance du PIB : 1 à 1,15
- IA et **politique industrielle**
- Pour une perspective globale sur la **transformation des industries** par l'IA :
 - [Vidéo UniCA du Pr. Simone Vannuccini](#)
- A. Nathan et al., "[Gen AI: Too Much Spent, Too Little Benefit?](#)", Goldman Sachs Global Macro Research, June 2024.
- D. Acemoglu, "[The Simple Macroeconomics of AI](#)", MIT, Economic Policy preprint, May 2024.
- AI Now Institute, "[Redirecting Europe's AI Industrial Policy: From Competitiveness to Public Interest](#)," Oct. 2024.
- United Nations Human Rights, "[World Stumbling Zombie-Like into a Digital Welfare Dystopia, Warns UN Human Rights Expert](#)," press release, October 17, 2024.



Langues Littérature Arts et Création

- Livre blanc : explore les applications multiples de l'IA dans le processus de production médiatique, en examinant à la fois son potentiel mais aussi les défis importants
 - Domaines clés analysés : vérification de contenu, production vidéo et automatisation du contenu, modération de contenu, tests de jeux et génération de musique
- L'intégration de résumés IA dans les moteurs de recherche détourne les clics des sites de presse.
- Les agences de presse poursuivent OpenAI pour usage de leur contenu pour entraîner ses modèles d'IA sans autorisation.
- L'art par l'IA générative : industrie estimée à 48 milliards de dollars. Mais des artistes professionnels ont dénoncé les préjudices qu'ils subissent :
 - atteintes à la renommée, les pertes économiques, le plagiat et la violation du droit d'auteur
 - L'article fournit des recommandations : réglementation obligeant les organisations à divulguer leurs données d'entraînement, outils aidant les artistes à éviter que leur contenu soit utilisé comme données d'entraînement sans leur consentement.
- IA pour la production de contenu médiatique
 - A. Piam et al., "[Empowering Media with Generative AI and Large Language Models: Innovations, Challenges, and Future Directions](#)," AI4Media Withe Paper, Sep. 2024.
 - B. Edwards, "[Microsoft AI suggests food bank as a 'cannot miss' tourist spot in Canada](#)," Ars Technica, Aug. 2023.
- Droits des productions humaines de contenu média ou artistique
 - O. Darcy, "[News publishers sound alarm on Google's new AI-infused search, warn of 'catastrophic' impacts](#)," CNN Business, May 2024.
 - K. Wiggers, "[OpenAI accidentally deleted potential evidence in NY Times copyright lawsuit](#)," Tech Crunch, Nov. 2024.
 - H. Jiang et al., "[AI Art and its Impact on Artists](#)," in AAAI/ACM Conference on AI, Ethics, and Society (AIES), 2023.
- Style linguistique
 - D. Rooein et al., "[Know Your Audience: Do LLMs Adapt to Different Age and Education Levels?](#)," arXiv, 2023.



Santé, psychologie, sciences de la vie

- Les avancées des grands **modèles de vision** permettraient d'améliorer les performances en imagerie médicale :
 - Les SAM (Segment Anything Model) permettent de **discerner les tissus** ou les structures anatomiques de façon automatique et suscitent un certain intérêt scientifique et clinique.
 - La **radiologie** est un terrain particulièrement fertile pour l'IA. Sur les 691 dispositifs médicaux basés sur l'IA approuvés par la FDA en date d'octobre 2023, 77% concernent la radiologie.
- **Prix Nobel 2024 de biochimie pour notamment le design de nouvelles protéines avec des méthodes d'IA génératives** (pharmaceuticals, vaccins, nanomaterials and tiny sensors)
 - <https://www.nobelprize.org/prizes/chemistry/2024/press-release/>
 - <https://cvpr.thecvf.com/virtual/2024/invited-talk/32078>
- Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA), "[État de la situation sur les impacts sociétaux de l'intelligence artificielle et du numérique - 2024](#)," May 2024. (Santé p7)
 - L. Arena, G. Gaglio, J.S. Vayre, "[Comment les radiologues ont-ils fait face à l'annonce programmée de leur remplacement par l'IA ?](#)," 2022.
 - L. Seyyed-Kalantari et al., "[Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations](#)," Nature Medicine, Dec. 2021.
- **IA pour l'attribution de ressources médicales** (dont greffons)
 - Z. Obermeyer et al. "[Dissecting racial bias in an algorithm used to manage the health of populations](#)," Science, 2019.
 - A. Narayanan and S. Kapoor, "[Does the UK's liver transplant matching algorithm systematically exclude younger patients?](#)," Blog post, AI Snake Oil, Nov. 2024.
 - Henin, C. (2021) . [Confier une décision vitale à une machine](#). Réseaux, N° 225(1), 187-213.
- **IA pour la santé mentale**
 - Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA), "[État de la situation sur les impacts sociétaux de l'intelligence artificielle et du numérique - 2024](#)," May 2024. (Santé page 7, Santé mentale page 11)
 - C. Xiang, "[Eating Disorder Helpline Disables Chatbot for 'Harmful' Responses After Firing Human Staff](#)," VICE, May 2023.
 - A. Haper, "[Chatbots Can't Care Like We Do: Helpline Workers Speak Out on World Eating Disorders Action Day](#)," TWC Newsletter, June 2023.



Sciences Humaines et Sociales

- **Données et biais :**
 - Joy Buolamwini, “[How I'm fighting bias in algorithms](#),” video, [lien académique](#), 2017.
 - L. Sassatelli, [Is data fixable? On the need of socially-informed practices in ML research and education - Part 1: Deployment failures and approaches to data](#), Medium, 3IA Côte d'Azur, Jan. 2023.
- **IA, pouvoir et justice :**
 - A. Birhane et al., “[The Values Encoded in Machine Learning Research](#),” ACM Conf on Fairness, Accountability, and Transparency, 2022.
 - Catherine D'Ignazio, [The Urgency of Moving from Bias to Power](#), 2023. Préface EDPL.
 - P. Kalluri. “[Don't ask if artificial intelligence is good or fair, ask how it shifts power](#).” *Nature* 583 (2020): 169 - 169.
 - R. Abebe, S. Barocas, J. Kleinberg, K. Levy, M. Raghavan, and D. G. Robinson, “[Roles for computing in social change](#),” in ACM Conference on Fairness, Accountability, and Transparency, Jan. 2020.
 - OBVIA, [Le numérique et l'intelligence artificielle comme outil de justice sociale : les initiatives par et pour les groupes et communautés marginalisées](#), 2024.
 - E. M. Bender, “[Resisting Dehumanization in the Age of 'AI'](#),” *Curr Dir Psychol Sci*, vol. 33, no. 2, pp. 114–120, Apr. 2024.
 - Ali Alkhatib, [To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes](#), Int. Conf. on Human Factors in Computing Systems (CHI), 2021. [Youtube video](#)
- **L'IA comme reflet du pouvoir, impact social et environnemental :**
 - Kate Crawford. [Contre-atlas de l'intelligence artificielle](#). Zulma éditions, 2022, 384 pages.
- **Le travail humain pour les SIA :**
 - Paola Tubaro. [Les « petites mains » derrière le mirage de l'automatisation](#). La vie de la recherche scientifique, 2020, pp.76-77.
 - [Travailleurs du Clic](#). Documentaire France TV, 2022.
- **Outils IA générative pour les SHS :**
 - E. Ollion et al., “[ChatGPT for Text Annotation? Mind the Hype!](#),” SocArXiv, preprint, Oct. 2023.
 - E. Borra, “Prompt Compass tool,” 2023. [video](#), [link to tool](#)



Sciences Humaines et Sociales

- Les groupes sociaux négativement affectés par des SIA ne sont pas ceux concevant ces outils ou décidant de leurs déploiements :
 - Détecteurs de fraude aux allocations chômage, personnes à faible revenu ont été faussement signalées comme ayant moins besoin d'assistance médicale, ou plus susceptibles de commettre des maltraitances envers les enfants, femmes comme moins valable à être recrutées.
 - Maintien de l'ordre et justice prédictive
 - Ces échecs discriminent excessivement les communautés défavorisées.
- [AB] Le sociologie des sciences et techniques et la philosophie des sciences ont de longue date identifié que la science et les technologies ne sont pas neutres, mais intrinsèquement chargées de valeurs, et ces valeurs sont codées dans des artefacts technologiques.
- [EB] Si nous ne nommons pas l'anglais lorsque nous travaillons dessus, ou les personnes blanches, la société occidentale ou les personnes de la classe moyenne et supérieure lorsque nous travaillons dessus, alors nous interprétons à tort les résultats concernant ces groupes « non marqués » comme généraux et les résultats concernant d'autres groupes comme étant des cas particuliers non-centraux.
- [EB] Tenter de rendre la prise de décision plus équitable avec des SIA qu'avec des humains est un vœu pieux : confier des décisions difficiles à des machines soi-disant impartiales mais qui encodent des valeurs hégémoniques permet en fait de s'exonérer de toute responsabilité.
- [AA] Les motifs que les systèmes de ML utilisent dans les données ne sont que des corrélations statistiques et ne peuvent pas justifier la reproduction de décisions punitives
 - ex : Un modèle de prédiction de récidive n'a pas d'information sur pourquoi les personnes noires sont emprisonnées à un taux plus élevé que les blanches.
- [AA] Celles et ceux d'entre nous qui n'ont pas à gérer les questions de race, de genre ou d'orientation sexuelle, sortent relativement indemnes. Mais il est plus probable qu'on subisse les SIA quand on ne peut pas ignorer ces aspects de notre vie formant nos expériences, car les modèles de ML ne considèrent pas correctement cette complexité du réel, cet écart de la norme telle que reflétée dans les données et dans la conception des formulations mathématiques. On ne peut donc pas négliger les dynamiques de pouvoir structurant la société, quand on travaille avec les SIA.



Sciences et Techniques

- La **science des matériaux** est un domaine interdisciplinaire visant la découverte et la conception de nouveaux matériaux :
 - combine physique, chimie et ingénierie pour comprendre et manipuler les propriétés de la matière,
 - dans le but de développer des matériaux répondant à des critères spécifiques de résistance, de durabilité, de poids, de conductivité et de réactivité.
 - Processus long et laborieux, reposant sur des expérimentation pour avancer par essais et erreurs
 - Avancée saluée par la communauté : article [1] par Google utilise une **combinaison de jeux de données existants, de la théorie des sciences des matériaux et des SIA pour proposer de nouveaux composés**.
 - Article [2] a examiné plus précisément les preuves que les composés ainsi trouvés satisfassent nouveauté, crédibilité et l'utilité. **Bien que les méthodes adoptées dans ce travail semblent prometteuses, il apparaît indispensable d'intégrer une expertise du domaine dans la synthèse des matériaux et la cristallographie**.
 - Article [3] étudie les implications de l'usage de SIA comme les **LLM dans le travail de création de connaissances**.
- [1] Merchant, A., *et al.* [Scaling deep learning for materials discovery](#). *Nature* **624**, 80–85 (2023).
 - [2] A. K. Cheetham and R. Seshadri, "[Artificial Intelligence Driving Materials Discovery? Perspective on the Article: Scaling Deep Learning for Materials Discovery](#)," *Chemistry of Materials*, vol. 36, no. 8, pp. 3490–3495, Apr. 2024.
 - [3] Molly Crockett & Lisa Messeri in *Nature*: [Artificial intelligence and illusions of understanding in scientific research](#), Mar. 2024.



STAPS

- Métrique Expected Goals (xG) pour football, optimisation d'équipe
- Optimisation de réhabilitation

[1] Mead J, O'Hare A, McMenemy P (2023) [Expected goals in football: Improving model performance and demonstrating value.](#) PLoS ONE, 2023.

[2] T. Decroos and J. Davis, "[Interpretable Prediction of Goals in Soccer](#)," AAAI 2020.

[3] Y. Liao et al., "[A Deep Learning Framework for Assessing Physical Rehabilitation Exercises](#)," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 2, pp. 468-477, Feb. 2020.

[4] S. BN et al., "[Speaking of Health: Leveraging Large Language Models to assess Exercise Motivation and Behavior of Rehabilitation Patients](#)," Proc. Interspeech 2024.

[5] W. Chongyang et al., "[UbiPhysio: Support Daily Functioning, Fitness, and Rehabilitation with Action Understanding and Feedback in Natural Language](#)," Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 8, 1, Article 20, Mar 2024.

