UTL :: Comprendre l'Intelligence Artificielle (IA) et ses impacts

Références Bibliographiques

Thierry Brouard

Laboratoire d'Informatique Fondamentale et Appliquée de Tours

Université de Tours

Septembre 2023 - Décembre 2023

References

[1] Isaac ASIMOV : Runaround. *Astounding Science Fiction*, pages 94–103, mars 1942.

[2] J. MARKHOFF : Technology; A Celebration of Isaac Asimov. *New York Times*, avril 1992.

[3] Stuart RUSSEL et Peter NORVIG : *Intelligence Artificielle – Une approche moderne*. Pearson France, 4 édition, novembre 2021.

[4] Olivier HOUDÉ : *L'Intelligence*. QUE SAIS JE, Paris, septembre 2021.

[5] John MCCARTHY, Marvin Lee MINSKY, N ROCHESTER et Claude Elwood SHANNON : A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Rapport technique, août 1955.

[6] John MCCARTHY : Recursive functions of symbolic expressions and their computation by machine, Part I. *Commun. ACM*, 3(4):184–195, avril 1960.

[7] A.M. Turing Award Winners in Alphabetical Order. "https://amturing.acm.org/alphabetical.cfm".

[8] Intelligence artificielle : définition et utilisation. "https://www.europarl.europa.eu/news/fr/headlines/society/20200827STO85804/intelligence-artificielle-definition-et-utilisation", septembre 2020.

[9] Intelligence Artificielle. "https://www.larousse.fr/encyclopedie/divers/intelligence_artificielle/187257".

[10] A. M. TURING : Computing Machinery and Intelligence. *Mind*, LIX(236):433–460, octobre 1950.

[11] Warren S. MCCULLOCH et Walter PITTS : A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5(4):115–133, décembre 1943.

[12] D.O. HEBB : *The Organization of Behavior - A Neuropsychological Theory*. John Wiley & Sons, New York, 1949.

[13] Alan NEWELL et H.A. SIMON : The Logic Theory Machine – A Complex Information Processing System. Tech Report P-868, Rand Corporation, California, juin 1956.

[14] Franck ROSENBLATT : The Perceptron - a perceiving and recognizing automaton. Research Report 85-460-1, Cornell Aeronautical Laboratory, Buffalo - NY, janvier 1957.

[15] H. GELERNTER : Realization of a Geometry Theorem Proving Machine. pages 273–282, Paris, juin 1959.

[16] Franck ROSENBLATT : Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Rapport technique VG-1196-G-8, Cornell Aeronautical Laboratory, march 1961.

[17] A. G. IVAKHNENKO et Valentin Grigor'evich LAPA : *Cybernetics and Forecasting Techniques*. American Elsevier Publishing Company, 1967. Google-Books-ID: rGFgAAAAMAAJ.

[18] Peter JACKSON : *Introduction to expert systems*. International computer science series. Addison-Wesley, Harlow, England ; Reading, Mass, 3rd ed édition, 1999.

[19] Bruce G. BUCHANAN et Edward Hance SHORTLIFFE : *Rule-based expert systems : the MYCIN experiments of the Stanford Heuristic Programming Project*. Reading, Mass. : Addison-Wesley, 1984.

[20] Kunihiko FUKUSHIMA : Cognitron: A self-organizing multilayered neural network. *Biol. Cybernetics*, 20(3-4):121–136, 1975.

[21] Kunihiko FUKUSHIMA : Neocognitron. *Scholarpedia*, 2(1):1717, 2007.

[22] Sven BEHNKE : *Hierarchical Neural Networks for Image Interpretation*, volume 2766 de *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[23] David SLADE : Black Mirror : Bandersnatch, 2018.

[24] J. V. NEUMANN : Zur Theorie der Gesellschaftsspiele. *Math. Ann.*, 100(1):295–320, décembre 1928.

[25] Alan NEWELL et Herbert A. SIMON : Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM*, 19(3):14, mars 1976.

[26] Rémi COULOM : Efficient Selectivity and Backup Operators in Monte-Carlo Tree Search. *In* H. Jaap van den HERIK, Paolo CIANCARINI, H. H. L. M. Jeroen DONKERS et Jaap van den HERIK, éditeurs : *Computers and games: 5th international conference, CG 2006, Turin, Italy, May 29-31, 2006 ; revised papers*, numéro 4630 de Lecture notes in computer science. Springer, Berlin Heidelberg, 2007. Meeting Name: CG.

[27] David SILVER, Thomas HUBERT, Julian SCHRITTWIESER, Ioannis ANTONOGLOU, Matthew LAI, Arthur GUEZ, Marc LANCTOT, Laurent SIFRE, Dharshan KUMARAN, Thore GRAEPEL, Timothy LILLICRAP, Karen SIMONYAN et Demis HASSABIS : A

general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419):1140–1144, décembre 2018.

[28] Caroline LEFER-PALOS : Le jour où Deep Blue a battu Garry Kasparov aux échecs, août 2020.

[29] George BOOLE : *An Investigation of the Laws of Thought : On Which Are Founded the Mathematical Theories of Logic and Probabilities*. Mathesis. Vrin, Paris, 1992.

[30] Gottlob FREGE : *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. 1879.

[31] Jean ROHMER : Comprendre l'intelligence artificielle symbolique. "http://theconversation.com/comprendre-lintelligence-artificielle-symbolique-104033", novembre 2018.

[32] John MCCARTHY : Programs with Common Sense. *In Semantic Information Processing*, pages 403–418. MIT Press, Cambridge:MA, 1968.

[33] Allen NEWELL : The knowledge level. *Artificial Intelligence*, 18(1):87–127, janvier 1982.

[34] Martine BLANC, Elsie CHARRON et Michel FREYSSENET : Le développement des systèmes-experts en entreprise. Rapport technique 35, novembre 1989.

[35] Louis GACÔGNE : *512 problèmes corrigés en Pascal, C++, Lisp, Prolog*. Ellipses, Paris, janvier 1996.

[36] Jacques PITRAT : *METACONNAISSANCE. Futur de l'intelligence artificielle*. Hermès Sciences Publications, août 1997.

[37] Robert K. LINDSAY, Bruce G. BUCHANAN, Edward A. FEIGENBAUM et Joshua LEDERBERG : *Applications of artificial intelligence for organic chemistry: the DENDRAL project*. McGraw-Hill advanced computer science series. McGraw-Hill International Book Co, New York, 1980.

[38] V. YU, Lawrence M. FAGAN, S. W. BENNETT, W. CLANCEY, A. C. SCOTT, J. HANNIGAN, R. BLUM, B. BUCHANAN et Stanley N. COHEN : An Evaluation of MYCIN's Advice. 2005.

[39] CLÉMENT et THONNAT : Pilotage de procédures de traitement d'images pour la description morphologique de galaxies. 1992.

[40] A. N. CAMPBELL, V. F. HOLLISTER, R. O. DUDA et P. E. HART : Recognition of a Hidden Mineral Deposit by an Artificial Intelligence Program. *Science*, 217(4563):

927–929, septembre 1982. Publisher: American Association for the Advancement of Science.

[41] Richard B. MCCAMMON : Prospector II: Towards a knowledge base for mineral deposits. *Math Geol*, 26(8):917–936, novembre 1994.

[42] Badar ul ISLAM : Comparison of conventional and modern load forecasting techniques based on artificial intelligence and expert systems. *IJCSI International Journal of Computer Science Issues*, Vol. 8, septembre 2011.

[43] Cédric VILLANI, Marc SCHOENAUER, Yann BONNET, Charly BERTHET, Anne-Charlotte CORNUT, François LEVIN et Bertrand RONDEPIERRE : *Donner un sens à l'intelligence artificielle*. Mission Villani sur l'intelligence artificielle, mars 2018.

[44] Explainable Artificial Intelligence. "https://www.darpa.mil/program/explainable-artificial-intelligence", mars 2023.

[45] Tic-Tac-Toe Analysis using Clojure (Part 2). "https://www.occasionalenthusiast.com/tag/tic-tac-toe/", mai 2015.

[46] Hawking : « L'intelligence artificielle pourrait mettre fin à l'humanité ». *Le Monde.fr*, décembre 2014.

[47] Giuseppe BONACCORSO : *Machine Learning Algorithms*. Packt Publishing, 2017.

[48] Alan Mathison TURING : Intelligent Machinery. Research Report 61/205, National Physical Laboratory, 1948.

[49] A. L. SAMUEL : Some studies in machine learning using the game of checkers. *IBM J. Res. & Dev.*, 3(3):210–229, 1959.

[50] Douglas C. SCHMIDT : Google data collection. https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf, août 2018.

[51] Ryan NAKASHIMA : AP Exclusive: Google tracks your movements, like it or not, août 2018.

[52] Comment je suis devenue invisible, 2016.

[53] Michael COX et David ELLSWORTH : Application-Controlled Demand Paging for Out-of-Core Visualization. pages 235–244, Phoenix, AZ (USA), octobre 1997. IEEE.

[54] Invisibles - Les travailleurs du clic, 2022.

[55] Billy PERRIGO : Exclusive: The \$2 Per Hour Workers Who Made ChatGPT Safer, janvier 2023.

[56] Kartikay GOYLE, Quin XIE et Vakul GOYLE : DataAssist: A Machine Learning Approach to Data Cleaning and Preparation. 2023. Publisher: arXiv Version Number: 2.

[57] WIKIPEDIA : List of datasets for machine-learning research. https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research.

[58] François-Xavier LEDUC : Opinion | La confiance dans l'IA passe par la souveraineté des données, novembre 2021. Section: Idées & Débats.

[59] Introduction to Data-Centric AI. https://dcai.csail.mit.edu/, 2023.

[60] Daochen ZHA, Zaid Pervaiz BHAT, Kwei-Herng LAI, Fan YANG, Zhimeng JIANG et Shaochen ZHONG : Data-centric Artificial Intelligence: A Survey, mars 2023.

[61] List of cognitive biases, mars 2023.

[62] David LESLIE, Anjali MAZUMDER, Aidan PEPPIN, Maria K WOLTERS et Alexa HAGERTY : Does "AI" stand for augmenting inequality in the era of covid-19 healthcare? *BMJ*, page n304, mars 2021.

[63] Google INC. : Responsible ai practices. https://ai.google/responsibility/responsible-ai-practices/, 2023.

[64] Cem DILMEGANI : Responsible AI: 4 Principles & Best Practices in 2023, avril 2022.

[65] Heinz HERRMANN : What's next for responsible artificial intelligence: a way forward through responsible innovation. *Heliyon*, 9(3):e14379, 2023.

[66] AI Fairness 360. https://github.com/Trusted-AI/AIF360.

[67] What-if Tool. https://pair-code.github.io/what-if-tool/index.html#about.

[68] Julia ANGWIN, Jeff LARSON, Surya MATTU et Lauren KIRCHNER : Machine Bias, mai 2016.

[69] Principle of least privilege. https://en.wikipedia.org/wiki/Principle_of_least_privilege, mars 2023.

[70] Matt FREDRIKSON, Somesh JHA et Thomas RISTENPART : Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures. *In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, pages 1322–1333, New York, NY, USA, octobre 2015. Association for Computing Machinery.

[71] Arvind NARAYANAN et Vitaly SHMATIKOV : Robust De-anonymization of Large Sparse Datasets. *In 2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 111–125, mai 2008. ISSN: 2375-1207.

[72] Arvind NARAYANAN et Vitaly SHMATIKOV : Robust de-anonymization of large sparse datasets: a decade later.

[73] Experimental Security Research of Tesla Autopilot. Tech Report, Tencent Keen Security Lab, mars 2019.

[74] Kevin EYKHOLT, Ivan EVTIMOV, Earlence FERNANDES, Bo LI, Amir RAHMATI, Chaowei XIAO, Atul PRAKASH, Tadayoshi KOHNO et Dawn SONG : Robust Physical-World Attacks on Deep Learning Models, avril 2018. arXiv:1707.08945 [cs].

[75] Jiawei SU, Danilo Vasconcellos VARGAS et Sakurai KOUICHI : One pixel attack for fooling deep neural networks. 2017. Publisher: arXiv Version Number: 7.

[76] Samuel G. FINLAYSON, John D. BOWERS, Joichi ITO, Jonathan L. ZITTRAIN, Andrew L. BEAM et Isaac S. KOHANE : Adversarial attacks on medical machine learning. *Science*, 363(6433):1287–1289, mars 2019. Publisher: American Association for the Advancement of Science.

[77] Chawin SITAWARIN, Arjun Nitin BHAGOJI, Arsalan MOSENIA, Mung CHIANG et Prateek MITTAL : DARTS: Deceiving Autonomous Cars with Toxic Signs. 2018. Publisher: arXiv Version Number: 3.

[78] Cathy O'NEIL et Sébastien MARTY : *Algorithmes, la bombe à retardement*. les Arènes, Paris, 2018.

[79] CAF : le numérique au service de l'exclusion et du harcèlement des plus précaires, octobre 2022. Section: Surveillance.

[80] Aaron SANKIN et Surya MATTU : Predictive Policing Software Terrible at Predicting Crimes. *Wired*, oct 2023. Section: tags.

[81] Parlement EUROPÉEN : Ai act. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence, 2023.

[82] Le Parlement européen adopte l'AI Act en plénière.

[83] PENGUIN, éditeur. *The Filter Bubble: What The Internet Is Hiding From You*. Eli Pariser, New York, NY, USA, 1er édition, mai 2011.

[84] David CHAVALARIAS : *Toxic data: comment les réseaux manipulent des opinions*. Flammarion, Paris, 2022.

[85] Giuliano da EMPOLI : *Les ingénieurs du chaos*. JC Lattès, Paris, mars 2019.

[86] Myanmar: L'atrocité des réseaux sociaux. Meta face au droit à réparations des Rohingyas (Synthèse). Rapport technique ASA 16/5933/2022, Amnesty International, London, England, septembre 2022.

[87] Joy BUOLAMWINI et Timnit GEBRU : Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *In Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, pages 77–91. PMLR, janvier 2018. ISSN: 2640-3498.

[88] Patrick GROTHER, Mei NGAN et Kayee HANAOKA : Face recognition vendor test part 3:: demographic effects. Rapport technique NIST IR 8280, National Institute of Standards and Technology, Gaithersburg, MD, décembre 2019.

[89] Joseph B. FULLER, Manjari RAMAN, Eva SAGE-GAVIN et Kristen HINES : Hidden Workers: Untapped Talent. Rapport technique, Harvard Business School, septembre 2021.

[90] Mathilde ROCHEFORT : Utilisé comme chatbot médical, GPT-3 conseille à un patient de se suicider, octobre 2020.

[91] Eloïse DUVAL : «Comme une drogue dans laquelle il se réfugiait» : ce que l'on sait du suicide d'un Belge ayant discuté avec une intelligence artificielle. Section: Médias.

[92] Pierre-Yves GERLAT : Oregon : les services de protection de l'enfance vont cesser d'utiliser un algorithme entraîné pour détecter la maltraitance des enfants. *ActuIA*, juin 2022.

[93] Walt Detmar MEURERS : Natural language processing and language learning. *The Encyclopedia of Applied Linguistics*, 2012.

[94] Yupeng CHANG, Xu WANG, Jindong WANG, Yuan WU, Linyi YANG, Kaijie ZHU, Hao CHEN, Xiaoyuan YI, Cunxiang WANG, Yidong WANG, Wei YE, Yue ZHANG, Yi CHANG, Philip S. YU, Qiang YANG et Xing XIE : A Survey on Evaluation of Large Language Models, octobre 2023. arXiv:2307.03109 [cs].

[95] Tomas MIKOLOV, Kai CHEN, Greg CORRADO et Jeffrey DEAN : Efficient estimation of word representations in vector space, 2013.

[96] Jeffrey PENNINGTON, Richard SOCHER et Christopher D. MANNING : GloVe: Global Vectors for Word Representation. *In Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543. 2014.

[97] S. SQUARTINI, A. HUSSAIN et F. PIAZZA : Preprocessing based solution for the vanishing gradient problem in recurrent neural networks. *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS '03.*, 5:V–V, 2003.

[98] Sepp HOCHREITER et Jürgen SCHMIDHUBER : Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, novembre 1997.

[99] G. V. HOUDT, Carlos MOSQUERA et G. NÁPOLES : A review on the long short-term memory model. *Artificial Intelligence Review*, pages 1–27, 2020.

[100] Ashish VASWANI, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan N. GOMEZ, Lukasz KAISER et Illia POLOSUKHIN : Attention Is All You Need, août 2023. arXiv:1706.03762 [cs].

[101] Jacob DEVLIN, Ming-Wei CHANG, Kenton LEE et Kristina TOUTANOVA : BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, mai 2019. arXiv:1810.04805 [cs].

[102] Alec RADFORD et Karthik NARASIMHAN : Improving language understanding by generative pre-training. 2018.

[103] Tom B. BROWN, Benjamin MANN, Nick RYDER, Melanie SUBBIAH, Jared KA-PLAN, Prafulla DHARIWAL, Arvind NEELAKANTAN, Pranav SHYAM, Girish SAS-TRY, Amanda ASKELL, Sandhini AGARWAL, Ariel HERBERT-VOSS, Gretchen KRUEGER, Tom HENIGHAN, Rewon CHILD, Aditya RAMESH, Daniel M. ZIEGLER, Jeffrey WU, Clemens WINTER, Christopher HESSE, Mark CHEN, Eric SIGLER, Mateusz LITWIN, Scott GRAY, Benjamin CHESS, Jack CLARK, Christopher BERNER, Sam MCCANDLISH, Alec RADFORD, Ilya SUTSKEVER et Dario AMODEI : Language Models are Few-Shot Learners, juillet 2020. arXiv:2005.14165 [cs].

[104] Alan D. THOMPSOM : Gpt-4. https://lifearchitect.ai/gpt-4/, 2023.

[105] Junjie YE, Xuanting CHEN, Nuo XU, Can ZU, Zekai SHAO, Shichun LIU, Yuhan CUI, Zeyang ZHOU, Chao GONG, Yang SHEN, Jie ZHOU, Siming CHEN, Tao GUI, Qi ZHANG et Xuanjing HUANG : A Comprehensive Capability Analysis of GPT-3 and GPT-3.5 Series Models. http://arxiv.org/abs/2303.10420, mars 2023. arXiv:2303.10420 [cs].

[106] Maximilian SCHREINER : GPT-4 architecture, datasets, costs and more leaked. https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked/, juillet 2023.

[107] Eleni ADAMOPOULOU et Lefteris MOUSSIADES : Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2:100006, décembre 2020.

[108] Joseph WEIZENBAUM : ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45, janvier 1966.

[109] Benj EDWARDS : AI-powered Bing Chat spills its secrets via prompt injection attack. https://arstechnica.com/information-technology/2023/02/ai-powered-bing-chat-spills-its-secrets-via-prompt-injection-attack/, février 2023.

[110] John SCHULMAN, Filip WOLSKI, Prafulla DHARIWAL, Alec RADFORD et Oleg KLIMOV : Proximal Policy Optimization Algorithms. http://arxiv.org/abs/1707.06347, août 2017. arXiv:1707.06347 [cs].

[111] Paul CHRISTIANO, Jan LEIKE, Tom B. BROWN, Miljan MARTIC, Shane LEGG et Dario AMODEI : Deep reinforcement learning from human preferences. http://arxiv.org/abs/1706.03741, juin 2017. arXiv:1706.03741 [cs, stat].

[112] Long OUYANG, Jeff WU, Xu JIANG, Diogo ALMEIDA, Carroll L. WAINWRIGHT, Pamela MISHKIN, Chong ZHANG, Sandhini AGARWAL, Katarina SLAMA, Alex RAY, John SCHULMAN, Jacob HILTON, Fraser KELTON, Luke MILLER, Maddie SIMENS, Amanda ASKELL, Peter WELINDER, Paul CHRISTIANO, Jan LEIKE et Ryan LOWE : Training language models to follow instructions with human feedback. http://arxiv.org/abs/2203.02155, mars 2022. arXiv:2203.02155 [cs].

[113] Jingfeng YANG, Hongye JIN, Ruixiang TANG, Xiaotian HAN, Qizhang FENG, Haoming JIANG, Bing YIN et Xia HU : Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. http://arxiv.org/abs/2304.13712, avril 2023. arXiv:2304.13712 [cs] version: 2.

[114] Alan D. THOMPSOM : Inside language models (from gpt-4 to palm). https://lifearchitect.ai/models/, novembre 2023.

[115] Our World in DATA : Computation used to train notable artificial intelligence systems. https://ourworldindata.org/grapher/artificial-intelligence-training-computation, 2023.

[116] Nastasia MICHAELS : 6,4 milliards de litres pour Microsoft : l'IA générative a-t-elle fait exploser la consommation d'eau des géants de la tech ?, septembre 2023. Section: Environnement.

[117] Josh TAYLOR : ChatGPT's alter ego, Dan: users jailbreak AI program to get around ethical safeguards. *The Guardian*, mars 2023.

[118] Sébastian SEIBT : L'intelligence artificielle ChatGPT et la démocratisation de la cybercriminalité. https://www.france24.com/fr/%C3%A9co-tech/20230116-l-intelligence-artificielle-chatgpt-et-la-d%C3%A9mocratisation-de-la-cybercriminalit%C3%A9, janvier 2023. Section: eco/tech.

[119] Eran SHIMONY et Omer TSARFATI : Chatting Our Way Into Creating a Polymorphic Malware. https://www.cyberark.com/resources/threat-research-blog/chatting-our-way-into-creating-a-polymorphic-malware, janvier 2023.

[120] Jack JONES : This has happened when a large website has let the AI write about Star Wars - Gearrice, juillet 2023. Section: Tech World.

[121] Farhad MANJOO : Opinion | ChatGPT Is Already Changing How I Do My Job. *The New York Times*, avril 2023.