

Conference Paper Title*

Elie Neghawi

*Electrical and Computer Engineering
Concordia University
Montreal, Canada
e_negh@encs.concordia.ca*

Yan Liu

*Electrical and Computer Engineering
Concordia University
Montreal, Canada
yan.liu@concordia.ca*

Abstract—Adoption of AI systems has been widely used across multiple industry domains at an alerting rate without the focus on its ethical concerns. To address those concerns, there are an increase number of AI ethics frameworks that have been suggested recently that focus on the algorithmic level rather on the systems level. Nonetheless, some of the system level approaches developed mostly cover a single level governance pattern of the system components in the entire software design lifecycle. However, the need to go beyond the single level system design AI ethics frameworks to allow not only a better responsible-AI-by-design, but also a trustworthy process patterns that abstract and link the underlying layers of responsible AI on each and every level. This paper illustrates a principal-to-practice guide of the multi-level governance within organizations across the globe for AI ethics frameworks. We outline three main areas of gap in organizations for AI ethics frameworks. Consecutively, we propose a multi-level governance pattern for responsible AI systems within organizations which is participatory, iterative, flexible and operationalizable that target those three main gap areas. Finally, to assist practitioners to apply the multi-level governance AI in organizations and the impact that it has on the industry level, we will translate into effective and responsible AI practices.

Index Terms—AI, AI ethics, trustworthy AI, AIM-LOps, AIOps, software engineering, software architecture, pattern, best practice

I. INTRODUCTION

Artificial Intelligence (AI) reshaped our lives, helped people make better predictions and take more informed and wise decisions. However, these high tech are still in there infancy, and there remains much promise for AI to promote innovation and address global challenges that people face.

On the other hand, ethical concerns and anxieties are fuelling around AI. There are lots of enquiries on the trustworthiness and adoption of AI systems, including concerns about exacerbating inequality, digital divide, climate change and market concentration.

II. EASE OF USE

A. Maintaining the Integrity of the Specifications

The IEEEtran class file is used to format your paper and style the text. All margins, column widths, line spaces,

and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

III. PREPARE YOUR PAPER BEFORE STYLING

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections III-A–III-E below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads—L^AT_EX will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m²” or “webers per square meter”, not “webers/m²”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
- Use a zero before decimal points: “0.25”, not “.25”. Use “cm³”, not “cc”).

C. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \quad (1)$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

D. L^AT_EX-Specific Advice

Please use “soft” (e.g., `\eqref{Eq}`) cross references instead of “hard” references (e.g., (1)). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don’t use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in L^AT_EX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you’ve discovered a new method of counting.

BIB_T_EX does not work by magic. It doesn’t get the bibliographic data from thin air but from .bib files. If you use BIB_T_EX to produce a bibliography you must send the .bib files.

L^AT_EX can’t read your mind. If you assign the same label to a subsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

L^AT_EX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it’s supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won’t be any anyway) and it might stop a wanted equation number in the surrounding equation.

E. Some Common Mistakes

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum μ_0 , and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
- In American English, commas, semicolons, periods, question and exclamation marks are located within

quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)

- A graph within a graph is an “inset”, not an “insert”. The word *alternately* is preferred to the word “alternately” (unless you really mean something that alternates).
- Do not use the word “essentially” to mean “approximately” or “effectively”.
- In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
- Do not confuse “imply” and “infer”.
- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the “et” in the Latin abbreviation “et al.”.
- The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [2].

F. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

G. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

H. Figures and Tables

a) *Positioning Figures and Tables*: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

TABLE I
TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	Subhead	Subhead
copy	More table copy ^a		

^aSample of a Table footnote.



Fig. 1. Example of a figure caption.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks ...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number citations consecutively within brackets [2]. The sentence punctuation follows the bracket [2]. Refer

simply to the reference number, as in [2]—do not use “Ref. [2]” or “reference [2]” except at the beginning of a sentence: “Reference [2] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [2]. Papers that have been accepted for publication should be cited as “in press” [2]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [2].

REFERENCES

- [1] E. Neghawi and Y. Liu, “Evaluation of parameter update effects in deep semi-supervised learning algorithms,” *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 351–360, 2020.
- [2] S. Laine and T. Aila, “Temporal ensembling for semi-supervised learning,” *CoRR*, vol. abs/1610.02242, 2016.
- [3] A. Tarvainen and H. Valpola, “Weight-averaged consistency targets improve semi-supervised deep learning results,” *CoRR*, vol. abs/1703.01780, 2017.
- [4] S. Hadjis, C. Zhang, I. Mitliagkas, and C. Ré, “Omnivore: An optimizer for multi-device deep learning on cpus and gpus,” *CoRR*, vol. abs/1606.04487, 2016.
- [5] Y. Lin, S. Han, H. Mao, Y. Wang, and W. J. Dally, “Deep gradient compression: Reducing the communication bandwidth for distributed training,” *CoRR*, vol. abs/1712.01887, 2017.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CoRR*, vol. abs/1512.03385, 2015.
- [7] G. Huang, Z. Liu, and K. Q. Weinberger, “Densely connected convolutional networks,” *CoRR*, vol. abs/1608.06993, 2016.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [9] O. Chapelle, B. Schölkopf, and A. Zien, Eds., *Semi-Supervised Learning*. The MIT Press, 2006. [Online]. Available: <http://dblp.uni-trier.de/db/books/collections/CSZ2006.html>
- [10] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. J. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Józefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. G. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. A. Tucker, V. Vanhoucke, V. Vasudevan, F. B. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “Tensorflow: Large-scale machine learning on heterogeneous distributed systems,” *CoRR*, vol. abs/1603.04467, 2016.
- [11] E. R. Sparks, A. Talwalkar, V. Smith, J. Kottalam, X. Pan, J. E. Gonzalez, M. J. Franklin, M. I. Jordan, and T. Kraska, “MLI: an API for distributed machine learning,” *CoRR*, vol. abs/1310.5426, 2013.
- [12] V. N. Vapnik, *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.
- [13] M. A. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, J. M. McCauley, M. J. Franklin, S. Shenker, and I. Stoica, “Fast and interactive analytics over hadoop data with spark,” *login Usenix Mag.*, vol. 37, 2012.
- [14] K. K. Pal and K. S. Sudeep, “Preprocessing for image classification by convolutional neural networks,” in *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2016, pp. 1778–1781.

- [15] K. P. Murphy, *Machine learning : a probabilistic perspective*. Cambridge, Mass. [u.a.]: MIT Press, 2013.
- [16] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, "Deep networks with stochastic depth," *CoRR*, vol. abs/1603.09382, 2016.
- [17] W. Shen, Z. Wei, S. Huang, B. Zhang, J. Fan, P. Zhao, and Q. Zhang, "Interpretable compositional convolutional neural networks," *CoRR*, vol. abs/2107.04474, 2021.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *CoRR*, vol. abs/1502.01852, 2015.
- [19] H. Mania, X. Pan, D. Papailiopoulos, B. Recht, K. Ramchandran, and M. I. Jordan, "Perturbed iterate analysis for asynchronous stochastic optimization," 2016.
- [20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol. abs/1512.00567, 2015.
- [21] T. Tieleman, G. Hinton *et al.*, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.
- [22] X. Gastaldi, "Shake-shake regularization," *CoRR*, vol. abs/1705.07485, 2017.
- [23] P. Goyal, P. Dollár, R. B. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He, "Accurate, large minibatch SGD: training imagenet in 1 hour," *CoRR*, vol. abs/1706.02677, 2017.
- [24] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015.
- [25] G. Kim, "Recent deep semi-supervised learning approaches and related works," *CoRR*, vol. abs/2106.11528, 2021.
- [26] F. Baldassarre and H. Azizpour, "Explainability techniques for graph convolutional networks," *CoRR*, vol. abs/1905.13686, 2019.
- [27] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," 2019.
- [28] R. Schwarzenberg, M. Hübner, D. Harbecke, C. Alt, and L. Henning, "Layerwise relevance visualization in convolutional text graph classifiers," *CoRR*, vol. abs/1909.10911, 2019.
- [29] P. E. Pope, S. Kolouri, M. Rostami, C. E. Martin, and H. Hoffmann, "Explainability methods for graph convolutional neural networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 10 764–10 773.
- [30] J. Chen, L. Song, M. J. Wainwright, and M. I. Jordan, "Learning to explain: An information-theoretic perspective on model interpretation," *CoRR*, vol. abs/1802.07814, 2018.
- [31] H. Yuan, J. Tang, X. Hu, and S. Ji, "XGNN: towards model-level explanations of graph neural networks," *CoRR*, vol. abs/2006.02587, 2020.
- [32] H. Yuan, H. Yu, S. Gui, and S. Ji, "Explainability in graph neural networks: A taxonomic survey," *CoRR*, vol. abs/2012.15445, 2020.
- [33] N. Xie, G. Ras, M. van Gerven, and D. Doran, "Explainable deep learning: A field guide for the uninitiated," *CoRR*, vol. abs/2004.14545, 2020.
- [34] L. M. Zintgraf, T. S. Cohen, T. Adel, and M. Welling, "Visualizing deep neural network decisions: Prediction difference analysis," *CoRR*, vol. abs/1702.04595, 2017.
- [35] D. Berthelot, N. Carlini, I. J. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," *CoRR*, vol. abs/1905.02249, 2019.
- [36] K. Sohn, D. Berthelot, C. Li, Z. Zhang, N. Carlini, E. D. Cubuk, A. Kurakin, H. Zhang, and C. Raffel, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," *CoRR*, vol. abs/2001.07685, 2020. [Online]. Available: <https://arxiv.org/abs/2001.07685>
- [37] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," 2017. [Online]. Available: <https://arxiv.org/abs/1703.01780>
- [38] M. Moradi and M. Samwald, "Post-hoc explanation of black-box classifiers using confident itemsets," *Expert Systems with Applications*, vol. 165, p. 113941, mar 2021. [Online]. Available: <https://doi.org/10.10162Fj.eswa.2020.113941>
- [39] A. Kessy, A. Lewin, and K. Strimmer, "Optimal whitening and decorrelation," *The American Statistician*, vol. 72, no. 4, pp. 309–314, jan 2018. [Online]. Available: <https://doi.org/10.1080%2F00031305.2016.1277159>
- [40] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015. [Online]. Available: <https://arxiv.org/abs/1512.00567>
- [41] W. Samek, T. Wiegand, and K. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," *CoRR*, vol. abs/1708.08296, 2017. [Online]. Available: <http://arxiv.org/abs/1708.08296>
- [42] M. Gaur, K. Faldu, and A. P. Sheth, "Semantics of the black-box: Can knowledge graphs help make deep learning systems more interpretable and explainable?" *CoRR*, vol. abs/2010.08660, 2020. [Online]. Available: <https://arxiv.org/abs/2010.08660>
- [43] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *CoRR*, vol. abs/1610.05492, 2016. [Online]. Available: <http://arxiv.org/abs/1610.05492>
- [44] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," *Proceedings of the National Academy of Sciences*, vol. 116, no. 44, pp. 22 071–22 080, oct 2019. [Online]. Available: <https://doi.org/10.1073%2Fpnas.1900654116>
- [45] Y. Zhang, P. Tino, A. Leonardis, and K. Tang, "A survey on neural network interpretability," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 5, pp. 726–742, oct 2021. [Online]. Available: <https://doi.org/10.1109%2FTetci.2021.3100641>
- [46] A. Preece, "Asking 'why' in ai: Explainability of intelligent systems – perspectives and challenges," *Int. J. Intell. Syst. Account. Financ. Manage.*, vol. 25, no. 2, p. 63–72, jun 2018. [Online]. Available: <https://doi.org/10.1002/isaf.1422>
- [47] S. R. Islam, W. Eberle, S. K. Ghafoor, and M. Ahmed, "Explainable artificial intelligence approaches: A survey," *CoRR*, vol. abs/2101.09429, 2021. [Online]. Available: <https://arxiv.org/abs/2101.09429>
- [48] M. Pocevič, G. Eilertsen, and C. Lundström, "Survey of XAI in digital pathology," in *Artificial Intelligence and Machine Learning for Digital Pathology*. Springer International Publishing, 2020, pp. 56–88. [Online]. Available: https://doi.org/10.1007%2F978-3-030-50402-1_4
- [49] J. Gerlings, A. Shollo, and I. D. Constantiou, "Reviewing the need for explainable artificial intelligence (xai)," *CoRR*, vol. abs/2012.01007, 2020. [Online]. Available: <https://arxiv.org/abs/2012.01007>
- [50] L. Querejeta Lomas, A. Goti Elordi, A. Almeida Escondrillas, and D. Lopez De Ipiña Gonzalez De Artaza, "A systematic literature review of artificial intelligence in fashion retail b2c," in *2021 6th International Conference on Smart and Sustainable Technologies (SpliTech)*, 2021, pp. 01–06.
- [51] M. Fox, D. Long, and D. Magazzini, "Explainable planning," *CoRR*, vol. abs/1709.10256, 2017. [Online]. Available: <http://arxiv.org/abs/1709.10256>
- [52] Y. Xie, S. Katariya, X. Tang, E. Huang, N. Rao, K. Subbian, and S. Ji, "Task-agnostic graph explanations," 2022. [Online]. Available: <https://arxiv.org/abs/2202.08335>
- [53] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *CoRR*, vol. abs/1609.02907, 2016. [Online]. Available: <http://arxiv.org/abs/1609.02907>
- [54] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, vol. 8, no. 8, 2019. [Online]. Available: <https://www.mdpi.com/2079-9292/8/8/832>
- [55] M. Sajjadi, M. Javanmardi, and T. Tasdizen, "Regularization with stochastic transformations and perturbations for deep semi-supervised learning," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ser. NIPS'16. Red Hook, NY, USA: Curran Associates Inc., 2016, p. 1171–1179.

- [56] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016. [Online]. Available: <https://proceedings.neurips.cc/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf>
- [57] J. Zhao, M. Mathieu, R. Goroshin, and Y. LeCun, "Stacked what-where auto-encoders," 2015. [Online]. Available: <https://arxiv.org/abs/1506.02351>