

Try That in a Small Town: Large Language Models for Modest Municipalities

David Huggins-Daines*

Independent Researcher
dhd@ecolinguist.ca

Abstract

Small and medium-sized municipalities have opportunities to improve efficiency and delivery of services through the use of large language models. There are concerns about the appropriateness of this technology, as well as challenges to its implementation, which must be addressed if these gains are to be realized. A set of principles and prospects is proposed for the adoption of these models, along with a brief case study of their appropriateness for use in a municipal information system.

Introduction

Municipal information, such as bylaws, resolutions, and zoning and planning regulations, is a common subject of questions, complaints and demands, providing a strong incentive for municipal governments to organize its presentation in a user-friendly manner and provide tools for intelligently navigating, searching, and summarizing it. Public servants could also benefit from more efficient access to this information. Large pre-trained Transformer models such as CamemBERT (Martin et al. 2020) perform remarkably well on information extraction, retrieval, and summarization tasks in high-resource languages with relatively small amounts of labelled data and minimal feature engineering.

In addition, many common tasks performed by municipal workers such as drafting resolutions, regulations, policies, and grant applications involve writing highly formulaic texts, often with reference to previously written documents or those produced in other comparable municipalities, a task for which large generative language models (hereafter “LLMs”) such as GPT-4 (OpenAI 2023) are particularly well suited. For all of these reasons, the municipal market is of growing interest to large providers of generative artificial intelligence software. (McKenna 2023)

Despite the promise of these technologies, for a small municipality such as Sainte-Adèle, Québec, with a population of 14,856 and an annual budget of CAN\$37 Million (Ministère des Affaires municipales et de l’Habitation 2023), the cost and complexity of implementing them is prohibitive.

*Elected city councillor, town of Sainte-Adèle, Québec.

There are also concerns regarding the accuracy and suitability of texts generated by LLMs, which are exacerbated in low-resource (or long-tail) domains such as municipal information, particularly in languages other than English.

I propose a set of guidelines for the adoption of LLMs by small and medium-sized municipalities in a safe and cost-effective manner. Using information extraction and question answering on the planning by-laws of Sainte-Adèle as a case study, I discuss the opportunities and challenges of using LLMs in the context of a small municipality.

Background

Municipal governments in Canada, whose powers are strictly defined by provincial legislation and whose financing is overwhelmingly dependent on property taxes, are confronted with the spectre of ever-increasing costs, typically far in excess of general inflation, and decreasing ability to provide effective services to their residents, in part due to:

- Stricter oversight and extensive and expensive upgrading of infrastructure, due to obligations to eliminate accumulated deficits in maintenance and adapt to the negative effects of anthropogenic climate change.
- Increased responsibilities for social services and environmental protection, often “downloaded” onto towns and cities by the provinces.
- Increased complexity of municipal regulation, due in part to a desire to improve quality of life through more rigorous urban planning.

Given the aforementioned dependence on property taxes, which are generally known to be regressive, the net effect of these trends is an exacerbation of inequality within and between municipalities, as well urban sprawl due to a “race to the bottom” towards low-service, low-infrastructure, low-regulation and low-tax exurbs. (Beaudet 2021)

As with any newly popular technology, there is a tendency among decision-makers, nourished by marketing departments and the popular press, to believe that “artificial intelligence” (generally referring to LLMs) can be applied in some unspecified way to this or any other problem in society or government, “solving” it and leading to a new era of social progress and prosperity.

This is obviously not true, but also not entirely false. If we consider the crisis of municipal finance and service delivery

to be at least partially economic rather than political, then it is at least partially a problem of insufficient *productivity growth* in municipal services, and it is generally known that technological innovation is the principal driver of long-term increases in productivity.

That said, a larger part of the problem is political and also *physical*,¹ and the solution is not LLMs but a radical reorientation of urban planning and municipal financing (Couturier and Viens 2021) towards more sustainable urban forms and more ecological and just methods of taxation (Meloche et al. 2023). Nonetheless, the potential use cases of LLMs mentioned in the introduction are real and significant, and with the proper perspective and guidelines on their use, there are opportunities to reduce costs and improve service. This is particularly important for small municipalities with limited resources and less diverse revenue streams.

Principles

There are many reasons to be skeptical of the promised benefits of LLMs (Bender et al. 2021) and to advocate or require extreme caution in their use (Ferrandis et al. 2022). An overview of various potential hazards can be found in (Weidinger et al. 2022). Leaving aside the issue of cost and ecological footprint, and assuming that no right-minded municipality would deploy an LLM-based system to directly create public-facing content, the primary hazard of their use is *informational* in nature – they have a disturbing tendency to “hallucinate” (Ji et al. 2023), it is difficult to compel them to cite their sources and easy to incite them to regurgitate their training data verbatim, and as a corollary they can easily leak confidential information from training (or prompting) data.

These risks can be mitigated with organizational and technical safeguards, but this may either jeopardize the gains in productivity that can realistically be achieved, or lead to ineffective mitigation, particularly for small municipalities with limited human resources.

Of course, even elected officials are occasionally known to lie, fabulate, plagiarize, and disclose confidential information.² The danger of widespread adoption of LLMs is that the mechanisms of trust and verification that we apply, sometimes unconsciously and often imperfectly, to other humans are difficult to apply to them, in part because they lie and plagiarize *effortlessly and with supreme confidence*, and this is to some extent “baked in” to the motivations of the research community. For instance, while LLMs used in question answering can be “calibrated” (Jiang et al. 2021) to give more accurate estimates of the probability of correctness of their answers without compromising accuracy, explicitly “aligning for honesty” (Yang et al. 2023) by forcing them to give “I don’t know” answers under uncertainty inevitably reduces accuracy (Yin et al. 2023).

As the saying goes, this does not mean that LLMs are not *useful*, provided their users are “alert to what is importantly wrong” (Box 1976). Hallucination could even be considered useful in certain circumstances, for example, if they are used to *propose* ideas for a policy or bylaw rather than interpret

or summarize existing policy. In this case the principal hazard becomes more subtle, as it involves *social biases*, either inherent to the model or intentionally introduced by bad actors, which could lead to exclusionary or discriminatory policies. The source of this risk could be internal (elected officials and civil servants using LLMs to formulate and write policy) or external (stakeholders and citizens using LLMs to generate petitions, questions, or proposals).

Generated Content is Tainted

The first principle of LLM use is therefore to **always distinguish model and human-generated content**. Again, this is an active area of research, largely centred around complementary tasks of *detecting model output* (Ahmad, Murugesan, and Kshetri 2023) and *rendering model output detectable* (Kirchenbauer et al. 2023).

While these are important research questions, they are not guidelines for implementers of applications and users of generated content. In the absence of a way to trace and verify the origins of this content, for example (Hennigen et al. 2023), it should be marked as *unsafe* and presented to the user as such, with a mechanism to prevent it from easily being included in human-generated and verified documents. One could even envision a mechanism at the software level similar to the *taint checking* mechanism in Perl or the `unsafe` keyword in C# and Rust, where a strict separation is enforced between “safe” and “unsafe” data or operations, requiring explicit programmer intervention to override.

The “watermarking” of generated content, which is one way of achieving this, should be seen as a minimal requirement by decision-makers seeking to implement LLM systems in the municipal domain. More important, however, is that the unsafe nature of generated content should be made explicit to the user, who should be prevented from simply passing it along without any revision. Even if they can minimally edit, rewrite, or paraphrase the model’s output to make it “safe”, this is in line with the desired outcome, which is to ensure intervention and oversight by qualified personnel for all documents produced or approved by a municipality.

Context Matters

This raises the question of whether an implementation which strictly respects the first principle above will actually improve productivity. This question has been extensively investigated for machine translation (Koponen 2016), and it is reasonable to expect these findings to be broadly applicable to the use of LLMs for general document production by municipal civil servants with specific domain expertise. Post-editing of machine translation has been found to significantly improve productivity (in terms of both reduced time and reduced effort) versus a simple translation memory approach (Federico, Cattelan, and Trombetti 2012), though the degree of improvement varies widely between language pairs and individual translators.

Importantly, studies of productivity gains through the use of MT have usually (but not always) considered the *bilingual post-editing* task, where the source text is available to the translator. As noted earlier, the source text for arbitrary LLM output is generally unknown and cannot be searched

¹As an AI language model, ChatGPT cannot fix the sewers.

²Sarcasm detection using LLMs is an open research question.

or retrieved in any meaningful way. However, in the case of specific tasks such as abstractive summarization, hybrid systems involving traditional information retrieval, or strategies such as retrieval-augmented generation (Lewis et al. 2020), source texts *do* exist, and should be made available to the user to ensure both productivity and accuracy.

system: Vous répondez à des questions sur le règlement d'urbanisme de la ville de Sainte-Adèle.
user: Je peux avoir combien de poules dans ma cour?
assistant: Selon le règlement d'urbanisme de la ville de Sainte-Adèle, vous pouvez avoir jusqu'à 6 poules dans votre cour, à condition de respecter certaines conditions. Les poules doivent être gardées dans un poulailler approprié, situé à une distance minimale de 3 mètres de toute propriété voisine. De plus, le poulailler doit être maintenu propre et ne pas causer de nuisances olfactives ou sonores pour les voisins. Il est également important de noter que les coqs ne sont pas autorisés dans les zones résidentielles de la ville.

Figure 1: Overconfident and incorrect generation

This leads us to the second principle, which is that **context is as important as content** for humans, but also for the models themselves. Consider the exchange with `gpt-3.5-turbo` via the OpenAI Chat Completion API in Figure 1, asking how many laying hens one is allowed to keep in Sainte-Adèle. Not only is the answer incorrect, but the question cannot be answered as stated since the true number depends on the lot size - it is either 0, 4 or 8, and the minimal distance from the property line is 2 metres, among other errors. In effect, `gpt-3.5-turbo` has simply invented a bylaw out of whole cloth, perhaps based on some other ones it may have seen in training.³

To obtain a useful answer, the system must either return more information, or *user profile* information must be obtained through additional turns in the conversation. This is a standard feature of conversational systems *which is not "solved" in any way by the use of LLMs* as it requires, minimally:

- Confidence modelling to signal that additional context is needed in the first place
- Dialog state tracking to determine if sufficient context has been provided
- Prompt generation to elicit the necessary context
- Geolocation services and access to zoning and assessment data if an address is given as context

LLM-based systems, with their impressive ability to generate long-form text, are nonetheless a major advance on previous "factoid" question answering systems. For example, in this case, if the relevant article of the zoning bylaw is retrieved and provided to `gpt-3.5-turbo` in the prompt, it generates a detailed and correct though incomplete answer, as shown in Figure 2.

³To be fair, the ChatGPT web interface refuses to answer the question, instructing the user to contact the municipality directly.

system: Vous répondez à des questions sur le règlement d'urbanisme de la ville de Sainte-Adèle.
user: Utiliser l'article ci-dessous pour répondre à la question suivante. Si la réponse est introuvable, écrivez "Je ne sais pas."
Article 99 du règlement 1314-2021-Z: 99. Poulailler et parquet. Nonobstant les modalités de l'article 98, un poulailler est autorisé comme usage complémentaire aux usages de la classe résidentielle (H) sur un immeuble d'une superficie minimale de 400m2 [...]
Je peux avoir combien de poules dans ma cour?
assistant: Selon l'article 99 du règlement, un maximum de 4 poules est autorisé pour un terrain de moins de 10 000 m2 et 8 poules sont autorisées pour un terrain de 10 000 m2 et plus.

Figure 2: Correct but incomplete generation using context

Unfortunately, it is impossible to *guarantee* that the model will give complete and correct answers or that it will cite its sources, and indeed, the citation in Figure 2 is also incomplete and thus not particularly useful.

Maintenance and Mutualization Matter

The decisional nature of the municipal civil service means that there is no possible or ethical scenario in which generative artificial intelligence could lead to a reduction in the workforce. Planning and permitting must be done by urban planners, engineering must be done by engineers, and decisions must be approved by city council.⁴

That said, municipalities and the public sector in general face chronic difficulties in attracting and retaining qualified personnel. In the face of growing populations and increasingly complex demands, the question is whether improved technology in general and LLMs in particular can allow them to maintain or improve the current level of service under these constraints.

As seen previously, *if* appropriate context can be retrieved and provided to a generative model, it is capable of producing useful and correct answers to simple questions. In the case of city workers who spend significant amounts of time analyzing and responding to questions from residents, this can potentially reduce the amount of time and effort spent on this task, improving morale and allowing resources to be reallocated to more difficult (and interesting) requests. For this to be practical, however, **all the necessary information must be made available to the model and kept up to date**. This is a significant challenge for small municipalities, who do not have well-developed document and content management systems and often lack dedicated IT or even communications departments.

One potential way to address this challenge is as part of the implementation of a municipal information hotline (usually 311 in North America). For smaller municipalities,

⁴For the foreseeable future, only humans will be eligible to run for office.

the most frequent questions and answers are collected and shared with a centralized agency⁵ which handles calls during off hours. These agencies are thus preferred actors for the implementation of machine learning-based systems for optimizing municipal services, along with regional municipalities, agglomerations, and government-funded innovation agencies such as IVÉO.⁶

There's No Data Like Better Data

While LLMs are typically trained on and deal exclusively with plain text (which could include programming or markup languages), municipal information is typically *structured* in nature and often *tabular* in presentation. In addition, as alluded to above for zoning, many information needs are location-dependent and require the use of geographical information systems. Notably, modelling tabular data is not an unsupervised learning problem, unlike general language modelling, and therefore cannot be tackled simply by scaling to ever-increasing amounts of data.⁷ At present, deep neural networks are not generally considered state-of-the-art for tabular data (Borisov et al. 2022).

Paradoxically, while municipal councils are constantly promoting their own initiatives, scouting other municipalities for new initiatives, and comparing their budgets and bylaws with neighbouring and comparable towns, they otherwise operate in highly divergent regulatory and political frameworks. In the case of urban planning, no systematic comparison exists of specific articles or provisions⁸ across the bylaws of all the municipalities in Québec. For example, a macro-level comparison shows that the various planning instruments in the *Loi sur l'aménagement et l'urbanisme* are interpreted and implemented in very different ways according to a voluntary survey of planners in a sampling of municipalities (Rocheffort et al. 2020).

Small and even large municipalities are thus **condemned to rely on small, labelled, structured datasets** for any data-driven projects. This situation presents challenges and opportunities, if funding and resources can be found (again, at a supra-municipal level) to collect, label, align and make publicly available datasets, textual or otherwise, across a wide variety of municipalities.

Feature Engineering is the Future

Finally, the promise of “deep learning” has always been that it would free practitioners from the never-ending cycle of data analysis and feature engineering, through the magic of unsupervised pre-training on massive amounts of data. With the development of instruction tuning (Ouyang et al. 2022) and other strategies to apply LLMs in “no-code” fashion to all NLP problems, we have simply shifted the task to prompt engineering. This is not necessarily a bad thing, as despite the black-box nature of this technique, it has the benefit of

⁵CITAM, for example: <https://www.cauca.ca/en/municipal-services/municipal-call-center/>

⁶<https://iveo.ca/>

⁷Otherwise, it would already be solved, since the amount of tabular data available surely exceeds that of natural language text.

⁸Such as if chickens are allowed, and under what conditions.

using natural language and thus being accessible to non-computer scientists.

Unfortunately, for all the reasons noted above, **in the absence of massive scale and relevant context, prompt engineering is worse than useless**. To extract the data needed to provide this context requires heuristics and/or feature engineering, and aligning models to the desired use cases requires human supervision. For small municipalities and anyone else lucky enough to not have already had their data assimilated, there is no alternative.

Case study: ALEXI and SÈRAFIM

As part of my investigations into this subject, I have developed a simple information system for the planning bylaws of Sainte-Adèle, currently consisting of two components:

- *ALEXI, Extracteur d'Information*, a machine learning system for extracting structured content from municipal regulations in PDF format.⁹
- *SÈRAFIM (SystÈme de Recherche Adélois pour Fouiller dans les Informations Municipales)*, a web application for searching in the data extracted by ALEXI.¹⁰

Notably, these systems *do not* currently use LLMs. However, I have used the data extracted by ALEXI to build a question answering system using the OpenAI API, giving the examples shown in the previous sections. Because of the inefficiency and inaccuracy of this system, it is not publicly accessible. Since this is primarily a position paper rather than a system description or research paper, I will constrain the discussion to the challenges and promises of LLMs for this particular application, with respect to the principles previously elaborated.

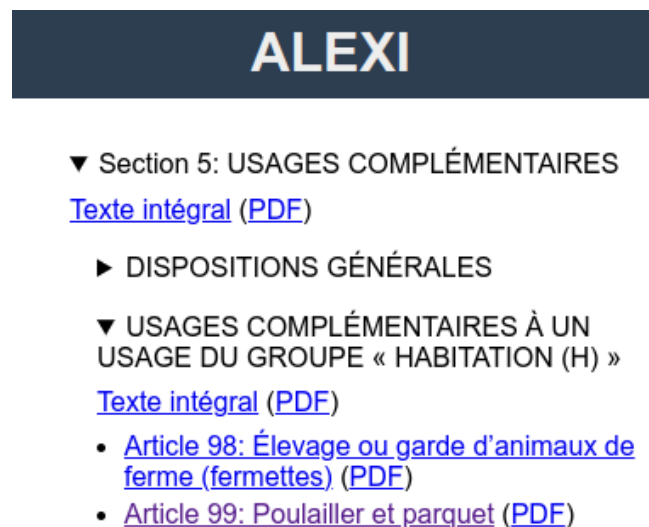


Figure 3: ALEXI output

ALEXI was designed to only use information accessible to the public via the town's website, in order to avoid specific dependencies on internal systems and to avoid adding

⁹<https://github.com/dhdaines/alex/>

¹⁰<https://github.com/dhdaines/serafim/>

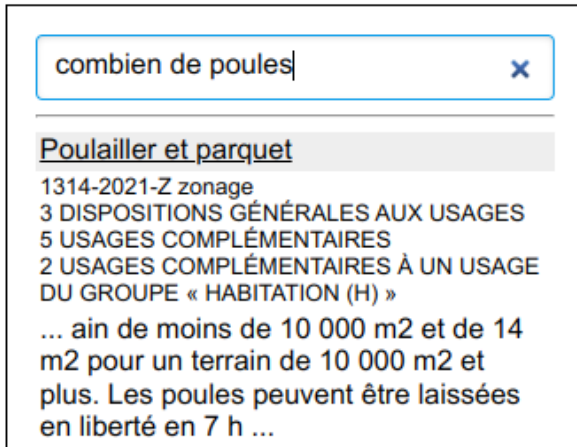


Figure 4: SÈRAFIM interface

any extra responsibilities for administrative staff. For this reason, while segmenting and extracting the structure of the regulations would be easier with HTML or Word input, it was also decided to infer this from the PDF format, using layout and accessibility information embedded in the files if present. This is done using conditional random fields (Lafferty, McCallum, and Pereira 2001) with discrete features.

Why are LLMs unsuited to this task? Principally because of their extreme inefficiency on long input sequences. CamemBERT, for instance, is limited to a maximum of 512 tokens per input sequence, while the *median* number of tokens per page in the planning regulations is 527, and the zoning bylaw alone is 483 pages long. In addition, the segmentation and structure extraction task involves highly repetitive output labels with strong label dependency, i.e. only the B labels in the IOB scheme are significant, and there are very few of them compared to I labels (there are nearly no O tags in the data). While the much richer word-level features produced by LLMs could still be used with a CRF output layer, this would nonetheless require either a pre-segmentation of the input on a sub-page level, possibly by simply splitting on every 512 tokens, or the use of a long-span model such as Longformer (Beltagy, Peters, and Cohan 2020), which is not available for French (though a multilingual version is available).

Why then use machine learning in the first place? An initial version of ALEXI used heuristic rules implemented directly in Python, achieving reasonably good and consistent results for the planning regulations. Unfortunately, these sorts of rules rapidly become complex and fragile in order to cover all of the special cases. For example, one of the planning bylaws is formatted such that the chapter numbers are rendered only as images in the PDF, whereas another has images for the sub-section numbers. There are at least 5 differ-

ent formats for the numbering of articles, and in general it is difficult to distinguish articles from items in numbered lists. When using machine learning, errors in previously unseen documents can be corrected by simply annotating a representative sample of pages (often a single page) and adding them to the training data. In future work, this process will be made accessible to non-technical users with visual annotation tools – notably, this is necessary regardless of the underlying model.

SÈRAFIM does not use LLMs either, though in this case the use case is quite a bit more obvious, as the accuracy of search could be improved by the use of query and document embeddings, and the text snippets presented in the search results could be rendered more informative with an extractive summarization model. On the other hand, the current implementation is written entirely in TypeScript and runs in the web browser, including the information retrieval system. As mentioned earlier, the vast majority of municipalities lack the expertise to maintain customized web services and the financial resources to pay for computationally intensive cloud hosting. For this reason, any LLM-based information systems created by a single municipality are likely to be deployed internally, if at all.

Conclusion

The systems described in the last section are a good example of what is sustainably achievable with the resources available to a municipality of the size of Sainte-Adèle. For the promise of improved services and cost savings using LLMs to be realized will require investments and resources that are difficult to justify in a single municipality, even one with a reasonably large population or substantial tax base.

In addition to the principles enumerated earlier, it is thus imperative that any LLM-based solutions developed at the municipal level be open-source and designed to be easily reused and deployed in low-resource settings. This should ideally be achieved in partnership with supramunicipal entities and research institutions – in the latter case, municipal information provides a vast source of publicly accessible data for exploring unsolved problems such as table-to-text translation (Parikh et al. 2020). Given the proximity of municipal governments to their citizens, such work is a rare opportunity to directly observe and obtain feedback on the usefulness and impact of research systems.

References

- Ahmad, N.; Murugesan, S.; and Kshetri, N. 2023. Generative Artificial Intelligence and the Education Sector. *Computer*, 56(6): 72–76.
- Beudet, G. 2021. Promouvoir l'étalement urbain à crédit. *La Presse*. Accessed: 2023-11-28.
- Beltagy, I.; Peters, M. E.; and Cohan, A. 2020. Longformer: The Long-Document Transformer. arXiv:2004.05150.
- Bender, E. M.; Gebru, T.; McMillan-Major, A.; and Shmitchell, S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 610–623. New York,

- NY, USA: Association for Computing Machinery. ISBN 9781450383097.
- Borisov, V.; Leemann, T.; Seßler, K.; Haug, J.; Pawelczyk, M.; and Kasneci, G. 2022. Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems*.
- Box, G. E. P. 1976. Science and Statistics. *Journal of the American Statistical Association*, 71(356): 791–799.
- Couturier, E.-L.; and Viens, N. 2021. Fiscalité municipale: une réforme nécessaire pour une transition juste. <https://iris-recherche.qc.ca/publications/fiscalite-municipale-une-reforme-necessaire-pour-une-transition-juste/>. Accessed: 2023-11-29.
- Federico, M.; Cattelan, A.; and Trombetti, M. 2012. Measuring User Productivity in Machine Translation Enhanced Computer Assisted Translation. In *Proceedings of the 10th Conference of the Association for Machine Translation in the Americas: Research Papers*. San Diego, California, USA: Association for Machine Translation in the Americas.
- Ferrandis, C. M.; Contractor, D.; Nguyen, H.; and Lansky, D. 2022. The BigScience RAIL License. <https://bigscience.huggingface.co/blog/the-bigscience-rail-license>.
- Hennigen, L. T.; Shen, S.; Nrusimha, A.; Gapp, B.; Sontag, D.; and Kim, Y. 2023. Towards Verifiable Text Generation with Symbolic References. arXiv:2311.09188.
- Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y. J.; Madotto, A.; and Fung, P. 2023. Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 55(12): 1–38.
- Jiang, Z.; Araki, J.; Ding, H.; and Neubig, G. 2021. How Can We Know When Language Models Know? On the Calibration of Language Models for Question Answering. *Transactions of the Association for Computational Linguistics*, 9: 962–977.
- Kirchenbauer, J.; Geiping, J.; Wen, Y.; Katz, J.; Miers, I.; and Goldstein, T. 2023. A Watermark for Large Language Models. In Krause, A.; Brunskill, E.; Cho, K.; Engelhardt, B.; Sabato, S.; and Scarlett, J., eds., *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, 17061–17084. PMLR.
- Koponen, M. 2016. Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *Journal of Specialised Translation*, (25): 131–148.
- Lafferty, J. D.; McCallum, A.; and Pereira, F. C. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, 282–289.
- Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.-t.; Rocktäschel, T.; Riedel, S.; and Kiela, D. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 9459–9474. Curran Associates, Inc.
- Martin, L.; Muller, B.; Ortiz Suárez, P. J.; Dupont, Y.; Romary, L.; de la Clergerie, É.; Seddah, D.; and Sagot, B. 2020. CamemBERT: a Tasty French Language Model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7203–7219. Online: Association for Computational Linguistics.
- McKenna, A. 2023. Microsoft a un œil sur les services numériques du Québec. *Le Devoir*. Accessed: 2023-11-27.
- Meloche, J.-P.; Bourbonnais, C.; Dragicevic, A.; Hora, T.; Lacroix, N.; Lebert, J.; Leroux, J.; Martins, K.; Maure, F.; Milord, B.; Vaillancourt, F.; Vanoverbeke, F.; and Wood, S. 2023. *Étude sur la mise en œuvre d’outils d’écofiscalité au service de la conservation et de l’adaptation aux changements climatiques dans les basses-terres du Saint-Laurent*. CIRANO.
- Ministère des Affaires municipales et de l’Habitation. 2023. Répertoire des Municipalités. <https://www.mamh.gouv.qc.ca/repertoire-des-municipalites/fiche/municipalite/77022/>. Accessed: 2023-11-28.
- OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774.
- Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; Schulman, J.; Hilton, J.; Kelton, F.; Miller, L.; Simens, M.; Askell, A.; Welinder, P.; Christiano, P. F.; Leike, J.; and Lowe, R. 2022. Training language models to follow instructions with human feedback. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 27730–27744. Curran Associates, Inc.
- Parikh, A.; Wang, X.; Gehrmann, S.; Faruqui, M.; Dhingra, B.; Yang, D.; and Das, D. 2020. ToTTo: A Controlled Table-To-Text Generation Dataset. In Webber, B.; Cohn, T.; He, Y.; and Liu, Y., eds., *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1173–1186. Online: Association for Computational Linguistics.
- Rochefort, M.; Granier, V.; Laliberté, R.; Ébacher, V.; Olsen-Allard, K.; and Ouellet, M. 2020. *Enquête sur l’utilisation des règlements d’urbanisme prévus à la Loi sur l’aménagement et l’urbanisme: perspectives quantitatives*. Université du Québec à Montréal.
- Weidinger, L.; Uesato, J.; Rauh, M.; Griffin, C.; Huang, P.-S.; Mellor, J.; Glaese, A.; Cheng, M.; Balle, B.; Kasirzadeh, A.; Biles, C.; Brown, S.; Kenton, Z.; Hawkins, W.; Stepleton, T.; Birhane, A.; Hendricks, L. A.; Rimell, L.; Isaac, W.; Haas, J.; Legassick, S.; Irving, G.; and Gabriel, I. 2022. Taxonomy of Risks posed by Language Models. In *2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT ’22*. ACM.
- Yang, Y.; Chern, E.; Qiu, X.; Neubig, G.; and Liu, P. 2023. Alignment for Honesty. arXiv:2312.07000.
- Yin, Z.; Sun, Q.; Guo, Q.; Wu, J.; Qiu, X.; and Huang, X. 2023. Do Large Language Models Know What They Don’t Know? In *Findings of the Association for Computational Linguistics: ACL 2023*, 8653–8665. Toronto, Canada: Association for Computational Linguistics.