Research Statement

Jonathan May

My work is in natural language processing (NLP), with significant contributions to the areas of machine translation [3; 5; 22–25; 29–32; 36; 44; 55; 56; 76], dialogue modeling [6–9; 12–16; 28; 38; 51], and information extraction [33; 35; 37; 39; 54; 66; 73; 75], as well as much cross-disciplinary work in collaboration with experts in Journalism [40; 57–62; 64; 65], Jewish Studies [20; 67], Law [63], Cryptography [1; 2], Online Behavioral Studies [6; 9; 11; 12; 28; 51], and Literary Criticism [17]. My research enables practical, usable software, resulting in dominant performance in DARPA LORELEI (2016–2022), IARPA MATERIAL (2017–2021), and DARPA LWLL (2019–2023) translation competitions, as well as Army transition efforts via SBIR. It is appealing to funders, yielding grants totaling nearly $70,000,000 over nine years. It is also appealing to the academic community, having won **Outstanding** and **Best Paper Awards** at **NAACL in 2018** (for analysis of the representational power of neural networks [10]), at **ACL in 2018** for a universal orthographic conversion tool [29], at **NAACL in 2022** (for examination of news article evolution [60]), and at **EMNLP in 2024** (for revelation of the differences between humans and language models in creative storytelling [68]).

My long-term goal in research is to enable effortless human communication across diverse domains despite intrinsic and extrinsic barriers such as language, disability, misunderstanding, and attack. While the challenges in reaching this goal are many, I view them principally as consequences of computational models' inability to learn while efficiently using resources, and much of my work operates through that lens. To address low-resource problems, my group employs combinations of parameter and hyperparameter optimization approaches, including metalearning [21; 46–50], parameter efficient training regimens [22], unsupervised learning [56], and data synthesis [53; 54]. We have pursued resource-efficient language model architectures that leverage inductive bias to reach the state of the art in numerous benchmarks [41–43]. We pioneered two transfer learning-based approaches [27; 76] that have since become major training paradigms.

To better understand the impact of modeling approaches in different domains, I have pursued many cross-disciplinary collaborations. For instance, we explored the role of language in games, learning models to navigate classic text-based video games [71; 72] and examining the role of language in the deceptive strategy game *Diplomacy* [26; 70]. We explored the degree to which language models can understand chemistry and food science well enough to predict flavors of chemical sequences [34]. We teamed with the Jewish Studies department at Cal State Northridge to address concerns about antisemitism in language models [67]. We applied machine translation technology to byte code decomposition [69]. And, I addressed the nature of explainability from a history of literature perspective [17]. In this statement, however, I will highlight two areas in my group's cross-disciplinary research that are particularly important because of their potential impact and role as foundations for future pursuits, namely, our (separate) efforts in **decipherment** and **bias**. Though seemingly disparate, these two research areas have a shared goal of **enabling communication**, a goal naturally shared by my extensive machine translation experience as well.

Decipherment. While barriers to communication due to language variation are examples of **incidental** communication blocks, we are also interested in enabling communication to avoid **intentional**, if not **adversarial** attempts to avoid communication success.[1] We considered the case of decipherment of historical encryptions. A multi-lingual, scalable solution to decipherment could unlock secrets in vast resources, such as the thousands of undeciphered documents in the Vatican Secret Archive [45]. We realized that, on the one hand, a naive example-based training regimen would not generalize to arbitrary cipher keys due to the strong assumptions baked into unit embeddings, and on the other hand, frequency-based tokenization could aid the historically difficult problem of multi-length cipher keys. Simple changes to training enabled us to build highly flexible multilingual substitution models with the new architecture [1] and the leveraging of byte-pair encoding yielded novel decipherment of a previously unknown historical text [2], a fragment of which is in Figure 1.

[1] We distinguish our work, which concerns historical decryption to unlock ancient secrets, from modern cryptography studies, which have more dual-use implications on privacy preservation.

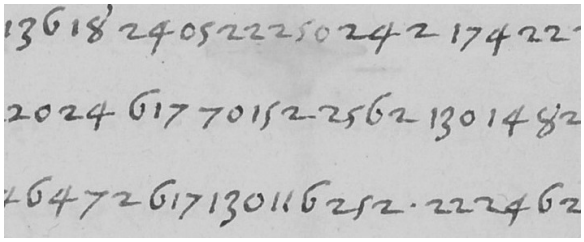

Figure 1. Fragment of the IA cipher, from the Vatican Secret Archives [45]. Our decipherment approaches [2] enabled its first complete decipherment.

Bias. Beyond explicit adversarial attacks such as encryption, the propagation of harmful stereotypes can inhibit the targets of those stereotypes from feeling free to express themselves, particularly in online communities. The effects of harmful stereotype propagation can be subtle and othering, manifesting when a language model has a high likelihood of generating a sentence expressing a harmful stereotype. Before we can combat harms, however, we need to understand what it means for a stereotype to be harmful. We recognized that, as our work will affect the community it applies to, it is incumbent on us to involve that community in our decision-making. In order to study anti-semitic and anti-Queer bias in language models, we diverged from the conventional top-down and didactic approach and instead designed **community-grounded** strategies to identify harmful bias. The approach we pioneered [19] was to solicit surveys and then convert free-text responses into formally structured datasets that allowed analysis by established methods [52]. We found that all of the dozens of models we tested had some bias against the Queer and Jewish communities, and that by exposing the models to data that was more representative of those communities, the bias was mitigated. We further found [20] that instruction models such as GPT-4 could not be trusted in these domains to do **qualitative coding**, a crucial part of dataset construction that determines the harms being expressed. These studies point to the essentiality of involving threatened communities when conducting research that affects them.

New Research Directions

Though challenges remain, our field has made great strides in the creation of models that can effortlessly understand and generate textual natural language. However, more than 95% of human language is produced and consumed without a textual medium. Further, we generally do not use language absent a rich experiential context that transcends language altogether. In 2020 I collaborated on a position paper [4] where we outlined the importance of this rich scope, including multisensory perception but also embodied behavior, proprioception, social relationship, and ultimately a reckoning of theory of mind, both at the individual and macro-societal level. As Big Tech has recognized the importance of modeling rich scope (e.g. in the evolution of self-driving car technology), they will pursue strategies that unite massive data and massive compute in furtherance of their own business agenda. However, if research using rich scope data is dominated by these interests, the resulting work will not be auditable or conducive to furthering scientific aims. I'm eager to apply the low-resource perspective that has underscored most of my lab's research to the challenge of developing world scope models for the purposes of better understanding the human learning process, without dependency on enormous computational costs.

My continued interest in disciplines that are not computer science is unabated, as I find it intoxicating to think about research from vastly different perspectives. Currently I am working with professors at American Jewish University to transcribe and translate diasporic languages, like Judeo-Persian, which are handwritten and exist only in fragments of text, in order to establish sociological relationships for anthropological purposes.[2] The goal of this effort is to create a Jewish Heritage Library (JHL)—a browsable, searchable historical map portal that invites the public to learn about Jewish cultures and engage with a massive, multilingual, international collection of documents, recordings, images, and databases. In order to make the JHL a reality, linkages must be found between the fragments. In principle one could use machine translation to accomplish this task. However, the corpora necessary for building typical machine translation models simply do not exist in sufficient volumes for these languages, so we cannot turn to traditional approaches. Diaspora languages are, however, closely related to one or more co-territorial or "parent" languages (e.g. Spanish for Ladino, French for Judeo-French; primarily German for Yiddish, though it's complicated), and a co-cultural component (Hebrew or Aramaic-derived terms for Jewish languages). Diaspora and co-territorial languages are generally somewhat mutually intelligible, but written

[2] https://www.jewishlanguages.org/jewish-language-project

forms vary in writing system and alphabet interpretation. For example, Judeo-Spanish (Ladino) uses the word bushkar (to search), where Spanish uses buskar. Because of this close relationship to larger-resource languages, I believe our aforementioned decipherment work [1], where we used a frequency-based conversion of substitution ciphers to crack encrypted ancient manuscripts, can be applied to uncover the mysteries behind the fragments.

I am analyzing legal decisions with professors from USC's law school to determine if bias can be detected in their anodyne prose, by inverting the methods we pioneered for removing [19; 20]. In that work, we found that many off-the-shelf language models had bias against the Queer and Jewish communities, and that by exposing the models to data that was more representative of those communities, the bias was mitigated. A question we are positing in this new effort is whether we can turn the experiment on its head, and instead expose a model to a selection of text from a source we believe to harbor bias, using the change in a model's resulting bias as an analytical tool to understand the source of the behavior-altering data.

While I have been a long-time fan of *Star Trek: The Next Generation* and the android Data in particular, I don't believe embodied, dialogue-capable agents, even when equipped with robotic human simulation infrastructure, will be adequate social replacements for humans in my lifetime. A shared human developmental experience is essential to forming friendship and trust bonds, and this will remain too hard to capture until we understand more about human learning itself. That said, limited human simulacra, even those as simple as a rubber duck, an ELIZA neo-Rogerian simulator, or more recently, GPT-o1, can prove useful for self-exploration, rehearsal for actual human-human conversation, and as inspiration for the creative process. Our early exploration into constructing SPOLIN, a chatbot that could participate as an improvisational comedy partner [15] was a creative inspiration that was used at comedy festivals[3] and remains a part of USC's undergraduate Improv for Engineers curriculum.[4] There is great potential in using models to brainstorm new ideas to pursue, or new directions of artistic expression, but we have qualitatively found that there is a limit to this utility, especially when we try to work with a model to produce humor, strong emotions, or other phenomena that seem to be correlated with **surprise**. Our recent analysis of LM story generation [68], which won an **Outstanding Paper Award** at EMNLP in 2024, analyzes the difference between generation by models and humans, but does not attempt to make model generation more human-like. We are working on surprisal-oriented objectives to try to imbue more creative spark into would-be creative partners.

I am fascinated by the connections between NLP and neuroscience, and the potential for analysis of neuronal data using computational techniques to better understand human language at the source, to improve the communicative abilities of those with physiological limitations, and to even enable new methods of human communication. The potential became apparent to me after learning about my colleagues' early successes in cross-lingual story identification [18]. Recent work [74] demonstrates the learnability of language-specific signal in neuronal recordings, and the viability of learning connections between the dense representations read in a scan and the ones we have engineered (e.g. GPT-4). However, many questions remain: To what degree is there still a disconnect between neuronal and 'siliconal' analysis approach? Are there non-invasive sensor techniques that can provide essential linguistic signal from the body? Is there a path forward to reliable generalized language recognition, such that we might one day be able to walk into a theater and don removable headgear in order to experience a new art form transmitted as direct neural stimulation? These are ambitious goals, and much as our advances over the past few decades have revealed gaps in our reckoning of societal, legal, and ethical implications, I anticipate struggles here as well. By anticipating the opportunities and challenges ahead, and by working cross-disciplinarily, as I have done as a founding member of USC's Center for Computational Language Studies, and will do as part of AIM, working with colleagues in Neuroscience, Linguistics, Art, and Public Policy, we can proceed boldly but not naively, and set the stage for a future of enhanced understanding of language and perhaps enhanced ability as well.

[3] https://www.youtube.com/watch?v=_QePlDFR09o

[4] https://magazine.viterbi.usc.edu/fall-2020/features/you-are-an-ai-yes-and-i-also-do-improv-comedy/

References

[1] Aldarrab, N., and **May, J.** Can sequence-to-sequence models crack substitution ciphers? In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language*

Processing (Volume 1: Long Papers) (Online, Aug. 2021), Association for Computational Linguistics, pp. 7226–7235.

[2] Aldarrab, N., and **May, J.** Segmenting numerical substitution ciphers. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (Abu Dhabi, United Arab Emirates, Dec. 2022), Association for Computational Linguistics, pp. 706–714.

[3] Barrón-Cedeño, A., Màrquez, L., Henríquez, Q. C. A., Formiga, L., Romero, E., and **May, J.** Identifying useful human correction feedback from an on-line machine translation service. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence* (2013), IJCAI '13, AAAI Press, p. 2057–2063.

[4] Bisk, Y., Holtzman, A., Thomason, J., Andreas, J., Bengio, Y., Chai, J., Lapata, M., Lazaridou, A., **May, J.**, Nisnevich, A., Pinto, N., and Turian, J. Experience grounds language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Online, Nov. 2020), Association for Computational Linguistics, pp. 8718–8735.

[5] Bremerman, J., Ren, X., and **May, J.** Machine translation robustness to natural asemantic variation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (Abu Dhabi, United Arab Emirates, Dec. 2022), Association for Computational Linguistics, pp. 3517–3532.

[6] Chang, R.-C., **May, J.**, and Lerman, K. Feedback loops and complex dynamics of harmful speech in online discussions. In *Social, Cultural, and Behavioral Modeling: 16th International Conference, SBP-BRiMS 2023, Pittsburgh, PA, USA, September 20–22, 2023, Proceedings* (Berlin, Heidelberg, 2023), Springer-Verlag, p. 85–94.

[7] Chawla, K., Lucas, G., **May, J.**, and Gratch, J. Opponent modeling in negotiation dialogues by related data adaptation. In *Findings of the Association for Computational Linguistics: NAACL 2022* (Seattle, United States, July 2022), Association for Computational Linguistics, pp. 661–674.

[8] Chawla, K., Ramirez, J., Clever, R., Lucas, G., **May, J.**, and Gratch, J. CaSiNo: A corpus of campsite negotiation dialogues for automatic negotiation systems. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Online, June 2021), Association for Computational Linguistics, pp. 3167–3185.

[9] Chen, K., He, Z., Chang, R.-C., **May, J.**, and Lerman, K. Anger breeds controversy: Analyzing controversy and¬†emotions on¬†reddit. In *Social, Cultural, and Behavioral Modeling* (Cham, 2023), R. Thomson, S. Al-khateeb, A. Burger, P. Park, and A. A. Pyke, Eds., Springer Nature Switzerland, pp. 44–53.

[10] Chen, Y., Gilroy, S., Maletti, A., **May, J.**, and Knight, K. Recurrent neural networks as weighted language recognizers. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (New Orleans, Louisiana, June 2018), Association for Computational Linguistics, pp. 2261–2271. **Outstanding Paper Award**.

[11] Cho, H., Gowda, T., Huang, Y., Lu, Z., Tong, T., and **May, J.** BotEval: Facilitating interactive human evaluation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)* (Bangkok, Thailand, Aug. 2024), Y. Cao, Y. Feng, and D. Xiong, Eds., Association for Computational Linguistics, pp. 107–116.

[12] Cho, H., Liu, S., Shi, T., Jain, D., Rizk, B., Huang, Y., Lu, Z., Wen, N., Gratch, J., Ferrara, E., and **May, J.** Can language model moderators improve the health of online discourse? In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (Mexico City, Mexico, June 2024), K. Duh, H. Gomez, and S. Bethard, Eds., Association for Computational Linguistics, pp. 7478–7496.

[13] Cho, H., Madotto, A., Lin, Z., Chandu, K., Kottur, S., Xu, J., **May, J.**, and Sankar, C. Continual dialogue state tracking via example-guided question answering. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (Singapore, Dec. 2023), H. Bouamor, J. Pino, and K. Bali, Eds., Association for Computational Linguistics, pp. 3873–3886.

[14] Cho, H., Sankar, C., Lin, C., Sadagopan, K., Shayandeh, S., Celikyilmaz, A., **May, J.**, and Beirami, A. Know thy strengths: Comprehensive dialogue state tracking diagnostics. In *Findings of the Association for Computational Linguistics: EMNLP 2022* (Abu Dhabi, United Arab Emirates, Dec. 2022), Association for Computational Linguistics, pp. 5345–5359.

[15] Cho, H., and **May, J.** Grounding conversations with improvised dialogues. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (Online, July 2020), Association for Computational Linguistics, pp. 2398–2413.

[16] Cho, H. J., Jedema, N. P., Ribeiro, L. F. R., Sharma, K., Szekely, P., Moschitti, A., Janssen, R., and **May, J.** Speechworthy instruction-tuned language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (Miami, Florida, USA, Nov. 2024), Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds., Association for Computational Linguistics, pp. 10652–10670.

[17] Dane, J. A., and **May, J.** *Begging The Question: Critical Reasoning in Chaucer Studies, Book History, and Humanistic Inquiry*. No. II in Mythodologies. Marymount Institute Press, 2019, ch. III.3: Evidence and Artificial Intelligence, pp. 295–208.

[18] Dehghani, M., Boghrati, R., Man, K., Hoover, J., Gimbel, S. I., Vaswani, A., Zevin, J. D., Immordino-Yang, M. H., Gordon, A. S., Damasio, A., and Kaplan, J. T. Decoding the neural representation of story meanings across languages. *Human Brain Mapping 38*, 12 (2017), 6096–6106.

[19] Felkner, V., Chang, H.-C. H., Jang, E., and **May, J.** WinoQueer: A community-in-the-loop benchmark for anti-LGBTQ+ bias in large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Toronto, Canada, July 2023), Association for Computational Linguistics, pp. 9126–9140.

[20] Felkner, V., Thompson, J., and **May, J**. GPT is not an annotator: The necessity of human annotation in fairness benchmark construction. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Bangkok, Thailand, Aug. 2024), L.-W. Ku, A. Martins, and V. Srikumar, Eds., Association for Computational Linguistics, pp. 14104–14115.

[21] Gheini, M., Ma, X., and **May, J.** Know where you're going: Meta-learning for parameter-efficient fine-tuning. In *Findings of the Association for Computational Linguistics: ACL 2023* (Toronto, Canada, July 2023), Association for Computational Linguistics, pp. 11602–11612.

[22] Gheini, M., Ren, X., and **May, J.** Cross-attention is all you need: Adapting pretrained Transformers for machine translation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (Online and Punta Cana, Dominican Republic, Nov. 2021), Association for Computational Linguistics, pp. 1754–1765.

[23] Gowda, T., and **May, J.** Finding the optimal vocabulary size for neural machine translation. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (Online, Nov. 2020), Association for Computational Linguistics, pp. 3955–3964.

[24] Gowda, T., You, W., Lignos, C., and **May, J.** Macro-average: Rare types are important too. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Online, June 2021), Association for Computational Linguistics, pp. 1138–1157.

[25] Gowda, T., Zhang, Z., Mattmann, C., and **May, J.** Many-to-English machine translation tools, data, and pretrained models. In *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations* (Online, Aug. 2021), Association for Computational Linguistics, pp. 306–316.

[26] Gu, F., Wongkamjan, W., Boyd-Graber, J. L., Kummerfeld, J. K., Peskoff, D., and **May, J.** Personalized help for optimizing low-skilled users' strategy, 2024.

[27] Hambardzumyan, K., Khachatrian, H., and **May, J.** WARP: Word-level Adversarial ReProgramming. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (Online, Aug. 2021), Association for Computational Linguistics, pp. 4921–4933.

[28] He, Z., **May, J.**, and Lerman, K. Cpl-novid: Context-aware prompt-based learning for norm violation detection in online communities. *Proceedings of the International AAAI Conference on Web and Social Media 18*, 1 (May 2024), 569–582.

[29] Hermjakob, U., **May**, **J.**, and Knight, K. Out-of-the-box universal romanization tool uroman. In *Proceedings of ACL 2018, System Demonstrations* (Melbourne, Australia, July 2018), Association for Computational Linguistics, pp. 13–18.

[30] Hermjakob, U., **May, J.**, Pust, M., and Knight, K. Translating a language you don't know in the chinese room. In *Proceedings of ACL 2018, System Demonstrations* (Melbourne, Australia, July 2018), Association for Computational Linguistics, pp. 62–67.

[31] Hopkins, M., and **May, J.** Tuning as ranking. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (Edinburgh, Scotland, UK., July 2011), Association for Computational Linguistics, pp. 1352–1362.

[32] Hopkins, M., and **May, J.** Models of translation competitions. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Sofia, Bulgaria, August 2013), Association for Computational Linguistics, pp. 1416–1424.

[33] Huang, L., Ji, H., and **May, J.** Cross-lingual multi-level adversarial transfer to enhance low-resource name tagging. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (Minneapolis, Minnesota, June 2019), Association for Computational Linguistics, pp. 3823–3833.

[34] Huang, T., Lee, D., Sweeney, J., Shi, J., Steliotes, E., Lange, M., May, J., and Chen, M. Foodpuzzle: Developing large language model agents as flavor scientists, 2024.

[35] Huang, X., **May, J.**, and Peng, N. What matters for neural cross-lingual named entity recognition: An empirical analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (Hong Kong, China, Nov. 2019), Association for Computational Linguistics, pp. 6394–6400.

[36] Jin, L., He, J., **May, J.**, and Ma, X. Challenges in context-aware neural machine translation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (Singapore, Dec. 2023), H. Bouamor, J. Pino, and K. Bali, Eds., Association for Computational Linguistics, pp. 15246–15263.

[37] Li, M., Zeng, Q., Lin, Y., Cho, K., Ji, H., **May, J.**, Chambers, N., and Voss, C. Connecting the dots: Event graph schema induction with path language modeling. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Online, Nov. 2020), Association for Computational Linguistics, pp. 684–695.

[38] Liu, S., Cho, H., Freedman, M., Ma, X., and **May, J.** RECAP: Retrieval-enhanced context-aware prefix encoder for personalized dialogue response generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Toronto, Canada, July 2023), Association for Computational Linguistics, pp. 8404–8419.

[39] Lu, D., Subburathinam, A., Ji, H., **May, J.**, Chang, S.-F., Sil, A., and Voss, C. Cross-lingual structure transfer for zero-resource event extraction. In *Proceedings of the 12th Language Resources and Evaluation Conference* (Marseille, France, May 2020), European Language Resources Association, pp. 1976–1981.

[40] Lu, M., Cho, H. J., Shi, W., **May, J.**, and Spangher, A. Newsinterview: a dataset and a playground to evaluate llms' ground gap via informational interviews, 2024.

[41] Ma, X., Kong, X., Wang, S., Zhou, C., **May, J.**, Ma, H., and Zettlemoyer, L. Luna: Linear unified nested attention. *CoRR abs/2106.01540* (2021).

[42] Ma, X., Yang, X., Xiong, W., Chen, B., Yu, L., Zhang, H., **May, J.**, Zettlemoyer, L., Levy, O., and Zhou, C. Megalodon: Efficient llm pretraining and inference with unlimited context length, 2024.

[43] Ma, X., Zhou, C., Kong, X., He, J., Gui, L., Neubig, G., **May, J.**, and Zettlemoyer, L. Mega: Moving average equipped gated attention. In *The Eleventh International Conference on Learning Representations* (2023).

[44] **May**, **J.**, Benjira, Y., and Echihabi, A. An arabizi-english

social media statistical machine translation system. In *Proceedings of the Eleventh Biennial Conference of the Association for Machine Translation in the Americas* (Vancouver, Canada, October 2014), Association for Machine Translation in the Americas.

[45] Megyesi, B., Esslinger, B., Fornés, A., Kopal, N., Láng, B., Lasry, G., de Leeuw, K., Pettersson, E., Wacker, A., and Waldispühl, M. Decryption of historical manuscripts: the decrypt project. *Cryptologia 44*, 6 (2020), 545–559.

[46] M'hamdi, M., Freedman, M., and **May, J.** Contextualized cross-lingual event trigger extraction with minimal resources. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)* (Hong Kong, China, Nov. 2019), Association for Computational Linguistics, pp. 656–665.

[47] M'hamdi, M., Kim, D. S., Dernoncourt, F., Bui, T., Ren, X., and **May, J.** X-METRA-ADA: Cross-lingual meta-transfer learning adaptation to natural language understanding and question answering. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Online, June 2021), Association for Computational Linguistics, pp. 3617–3632.

[48] M'hamdi, M., Ren, X., and **May, J.** Cross-lingual continual learning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Toronto, Canada, July 2023), Association for Computational Linguistics, pp. 3908–3943.

[49] M'hamdi, M., and **May, J.** Leitner-guided memory replay for cross-lingual continual learning. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (Mexico City, Mexico, June 2024), K. Duh, H. Gomez, and S. Bethard, Eds., Association for Computational Linguistics, pp. 7808–7821.

[50] M'hamdi, M., **May, J.**, Dernoncourt, F., Bui, T., and Yoon, S. Multilingual meta-distillation alignment for semantic retrieval. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2024), SIGIR '24, Association for Computing Machinery, p. 597–607.

[51] Moon, J., Lee, D.-H., Cho, H., Jin, W., Park, C., Kim, M., **May, J.**, Pujara, J., and Park, S. Analyzing norm violations in live-stream chat. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (Singapore, Dec. 2023), H. Bouamor, J. Pino, and K. Bali, Eds., Association for Computational Linguistics, pp. 852–868.

[52] Nangia, N., Vania, C., Bhalerao, R., and Bowman, S. R. CrowS-pairs: A challenge dataset for measuring social biases in masked language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Online, Nov. 2020), B. Webber, T. Cohn, Y. He, and Y. Liu, Eds., Association for Computational Linguistics, pp. 1953–1967.

[53] Pan, X., Gowda, T., Ji, H., **May, J.**, and Miller, S. Cross-lingual joint entity and word embedding to improve entity linking and parallel sentence mining. In *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)* (Hong Kong, China, Nov. 2019), Association for Computational Linguistics, pp. 56–66.

[54] Pan, X., Zhang, B., **May, J.**, Nothman, J., Knight, K., and Ji, H. Cross-lingual name tagging and linking for 282 languages. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vancouver, Canada, July 2017), Association for Computational Linguistics, pp. 1946–1958.

[55] Pighin, D., Màrquez, L., and **May, J.** An analysis (and an annotated corpus) of user responses to machine translation output. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)* (Istanbul, Turkey, may 2012), N. C. C. Chair), K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, J. Odijk, and S. Piperidis, Eds., European Language Resources Association (ELRA).

[56] Pourdamghani, N., Aldarrab, N., Ghazvininejad, M., Knight, K., and **May, J.** Translating translationese: A two-step approach to unsupervised machine translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Florence, Italy, July 2019), Association for Computational Linguistics, pp. 3057–3062.

[57] Spangher, A., Peng, N., Ferrara, E., and **May, J.** Identifying informational sources in news articles. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (Singapore, Dec. 2023), H. Bouamor, J. Pino, and K. Bali, Eds., Association for Computational Linguistics, pp. 3626–3639.

[58] Spangher, A., Peng, N., **May, J.**, and Ferrara, E. "don't quote me on that": Finding mixtures of sources in news articles. In *Proc. Computation+Journalism Symposium* (Online, March 2020).

[59] Spangher, A., Peng, N., **May, J.**, and Ferrara, E. Modeling "newsworthiness" for lead-generation across corpora. *CoRR abs/2104.09653* (2021).

[60] Spangher, A., Ren, X., **May, J.**, and Peng, N. NewsEdits: A news article revision dataset and a novel document-level reasoning challenge. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Seattle, United States, July 2022), Association for Computational Linguistics, pp. "127–157". **Outstanding Paper Award**.

[61] Spangher, A., **May, J.**, Shiang, S.-R., and Deng, L. Multi-task semi-supervised learning for class-imbalanced discourse classification. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (Online and Punta Cana, Dominican Republic, Nov. 2021), Association for Computational Linguistics, pp. 498–517.

[62] Spangher, A., Tumgoren, S., Welsh, B., Peng, N., Ferrara, E., and **May, J.** Tracking the newsworthiness of public documents. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Bangkok, Thailand, Aug. 2024), L.-W. Ku, A. Martins, and V. Srikumar, Eds., Association for Computational Linguistics, pp. 14150–14168.

[63] Spangher, A., Xue, Z., Wu, T.-L., Hansen, M., and **May, J.** LegalDiscourse: Interpreting when laws apply and to whom. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (Mexico City, Mexico, June 2024), K. Duh, H. Gomez, and S. Bethard, Eds.,

Association for Computational Linguistics, pp. 8536–8559.

[64] Spangher, A., Youn, J., DeButts, M., Peng, N., Ferrara, E., and **May, J.** Explaining mixtures of sources in news articles. In *Findings of the Association for Computational Linguistics: EMNLP 2024* (Miami, Florida, USA, Nov. 2024), Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds., Association for Computational Linguistics, pp. 15837–15859.

[65] Spangher, A., Youn, J., **May, J.**, and Peng, N. First steps towards a source recommendation engine: Investigating how sources are used in news articles. In *Proc. The Joint Computation + Journalism European Data & Computational Journalism Conference* (Zurich, Switzerland, June 2023).

[66] Subburathinam, A., Lu, D., Ji, H., **May, J.**, Chang, S.-F., Sil, A., and Voss, C. Cross-lingual structure transfer for relation and event extraction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (Hong Kong, China, Nov. 2019), Association for Computational Linguistics, pp. 313–325.

[67] **May, J.**, Felkner, V. K., and Thompson, J. We can have AI without antisemitism—if we want it. *AJS Perspectives* (Summer 2024), 36–38.

[68] Tian, Y., Huang, T., Liu, M., Jiang, D., Spangher, A., Chen, M., **May, J.**, and Peng, N. Are large language models capable of generating human-level narratives? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (Miami, Florida, USA, Nov. 2024), Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds., Association for Computational Linguistics, pp. "17659–17681". **Outstanding Paper Award**.

[69] Weideman, N., Felkner, V. K., Wu, W.-C., **May, J.**, Hauser, C., and Garcia, L. Perfume: Programmatic extraction and refinement for usability of mathematical expression. In *Proceedings of the 2021 Research on Offensive and Defensive Techniques in the Context of Man At The End (MATE) Attacks* (New York, NY, USA, 2021), Checkmate '21, Association for Computing Machinery, p. 59–69.

[70] Wongkamjan, W., Gu, F., Wang, Y., Hermjakob, U., **May, J.**, Stewart, B., Kummerfeld, J., Peskoff, D., and Boyd-Graber, J. More victories, less cooperation: Assessing cicero's diplomacy play. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Bangkok, Thailand, Aug. 2024), L.-W. Ku, A. Martins, and V. Srikumar, Eds., Association for Computational Linguistics, pp. 12423–12441.

[71] Yin, X., and **May, J.** Comprehensible context-driven text game playing. In *Proc. 2019 IEEE Conference on Games (CoG)* (August 2019), pp. 1–8.

[72] Yin, X., Weischedel, R., and **May, J.** Learning to generalize for sequential decision making. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (Online, Nov. 2020), Association for Computational Linguistics, pp. 3046–3063.

[73] Yu, P., **May, J.**, and Ji, H. Bridging the gap between native text and translated text through adversarial learning: A case study on cross-lingual event extraction. In *Findings of the Association for Computational Linguistics: EACL 2023* (Dubrovnik, Croatia, May 2023), Association for Computational Linguistics, pp. 754–769.

[74] Zada, Z., Goldstein, A., Michelmann, S., Simony, E., Price, A., Hasenfratz, L., Barham, E., Zadbood, A., Doyle, W., Friedman, D., Dugan, P., Melloni, L., Devore, S., Flinker, A., Devinsky, O., Nastase, S. A., and Hasson, U. A shared model-based linguistic space for transmitting our thoughts from brain to brain in natural conversations. *Neuron 112*, 18 (Sept. 2024), 3211–3222.e5. Publisher: Elsevier.

[75] Zhang, B., Lin, Y., Pan, X., Lu, D., **May, J.**, Knight, K., and Ji, H. Elisa-edl: A cross-lingual entity extraction, linking and localization system. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (New Orleans, Louisiana, June 2018), Association for Computational Linguistics, pp. 41–45.

[76] Zoph, B., Yuret, D., **May, J.**, and Knight, K. Transfer learning for low-resource neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (Austin, Texas, November 2016), Association for Computational Linguistics, pp. 1568–1575.