Research Statement

JONATHAN MAY

Major Research Accomplishments

My research goal is to enable effortless human communication across myriad roadblocks, be they language difference, disability, misunderstanding, or adversarial attack. I have done this chiefly through the broad research area of *natural language processing* (NLP). My early career focused on enabling understanding and reducing information overload, with extensive work on machine translation [4; 5; 21–24; 28–31; 35; 42; 47; 48; 57], information extraction [32; 34; 36; 39; 46; 52; 55; 56], and dialogue modeling [6–9; 12–16; 27; 37; 44].

My scholarship has been recognized with four major awards that span theoretical formalisms, practical tooling, and structural analysis. In 2018, we received an **Outstanding Paper Award** at NAACL for establishing the theoretical equivalence between recurrent neural networks and weighted finite automata [10], bridging the gap between deep learning and formal language theory. That same year, we won the **Best Demo Award** at ACL for 'uroman' [29], a universal romanization tool now ubiquitous in multilingual NLP. Our work on modeling news revision histories earned an **Outstanding Paper Award** at NAACL 2022 [50]. Most recently, our focus on the limitations of generative models in creative domains led to an **Outstanding Paper Award** at EMNLP 2024 [53], where we demonstrated the critical structural and affective gaps between AI and human storytelling.

My research has attracted more than $70 million in funding over ten years from government sources such as DARPA, IARPA, and ARPA-H, and from private entities such as The Physicians' Foundation, Amazon, and Open Philanthropy (now Coefficient Giving). I am strongly collaborative in nature, having launched funded efforts with UMichigan, UIUC, USydney, MIT, and UMD, among others. I also prize cross-disciplinary initiatives, which have resulted in collaboration with colleagues in fields such as Religious Studies [20], Bibliography [18], Journalism [51], and Food Science [33].

Ongoing and Future Research Goals

My more recent work operates through a new lens: moving from decoding *language* to decoding *intent*. We have entered an era where generating fluent text is no longer the bottleneck. The new challenge is that models, while fluent, often fail to capture the pragmatic nuance of human interaction. To address this, my research now focuses on **communicative efficacy**: building systems that understand narrative structure, negotiate strategy, detect subtle bias, and align with human stylistic intent. I am eager to bring this focus to AIM, where the convergence of computer science, social science, and public policy offers the ideal environment for my next phase of work: creating computational partners that truly collaborate with humans.

Efficiency and Robustness in High-Stakes Environments. My group has a proven track record of building robust systems for resource-constrained environments. We operate on the principle that computational models must learn efficiently to be practically useful. We pioneered transfer learning approaches that have become standard training paradigms [26; 57] and developed adaptation strategies [21] that drastically reduce the computational cost of fine-tuning large pre-trained models. So that these models perform in the "long tail" of data distributions, we have also designed architectures and evaluation objectives that leverage inductive bias to reach state-of-the-art performance, even on rare types [23; 40; 41].

This theoretical rigor translates into usable software. My team achieved dominant performance in high-stakes government evaluations, including the DARPA LORELEI (2016-2022) and IARPA MATERIAL (2017-2021) programs [29; 34; 43; 46]. We have also applied efficient learning techniques to historical decipherment, successfully unlocking previously unknown ciphers from the Vatican Secret Archives [1; 2]. This dual foundation of algorithmic efficiency and system-building rigor ensures that while my future work will explore high-level semantics and pragmatics, it will remain grounded in engineering reality and rigorous evaluation.

Controlling Generation for Style, Narrative, and Framing. Building on this technical foundation, my lab is now addressing the fact that current language models often default to a generic, "anodyne" style that fails to

engage users or convey subtext. To fix this, we are moving beyond simple generation to engage deeply with the style change problem, developing methods to control *how* a model speaks, not just *what* it says.

True communication requires understanding structure. We first addressed this in the news domain by developing datasets and challenges to model document-level revision histories [50]. We have since extended this lens to creative storytelling. While our early work on improvisational agents demonstrated the potential for open-ended computational creativity [15], our recent analysis reveals that even modern systems struggle to maintain the surprisal and pacing required for human-level engagement, as shown in Figure 1. This analysis, which won an **Outstanding Paper Award** at EMNLP 2024, used a framework of **macro**-level story arcs, **meso**-level turning points, and **micro**-level affective dimensions, to identify critical gaps in how models manage narrative tension [53].

To address these gaps, we are developing fine-grained control over output style to better align with user intent. For individual personalization, we introduced **TICL** (Trial-Error-Explain In-Context Learning), a tuning-free method that adapts a model to a user's unique **voice** using fewer than ten examples, bypassing the need for expensive parameter updates [17]. Complementing this, our STAMP system re-purposes iterative optimization techniques from our statistical machine translation past [11; 30; 45] to balance the conflicting objectives of style modification and semantic fidelity [38]. This trajectory defines my approach to creating personalized systems that can truly adapt to a user's cognitive and emotional context.
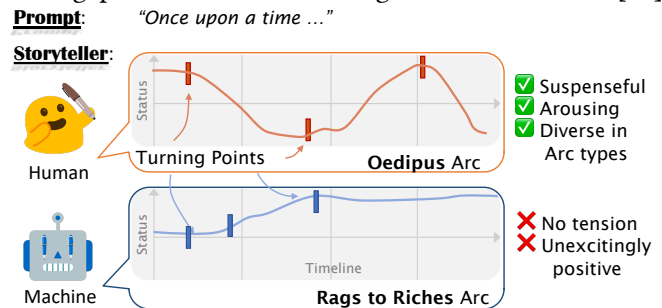

Figure 1. Visualizing the "Cognitive Gap" in Narrative. Our analysis [53] reveals that models fail to replicate the structural tension inherent to human storytelling, shying away from negativity and complex arcs. Addressing this divergence is central to my research on **Cognitive Alignment**—moving systems from surface-level imitation to deep structural resonance with human thought.

Interdisciplinary Impact: AI as a Collaborative Partner. My work increasingly bridges the gap between computer science and the social sciences, aiming to quantify and improve how automated systems affect human behavior, and to transform automated systems from static tools into dynamic teachers.

Interactive Learning and Pedagogical Partners. In earlier work we explored negotiation dynamics, releasing the **CaSiNo** annotated negotiation corpus [8] as the first instructive large-scale record of how negotiations take place. More recently, through the DARPA-funded **ALLAN** and **CHIRON** projects, I have collaborated with UMD's Jordan Boyd-Graber to explore how agents can scaffold human strategic thinking. Our recent work analyzing *Cicero* (Meta's Diplomacy agent) revealed a critical pedagogical gap: while the model could devise superior strategies, it failed to communicate them in a way that engendered human trust or comprehension [54]. We found, however, that when properly framed, strategic advice from Cicero helps novice players perform closer to expert levels, demonstrating the potential of strategy engines as effective tutors in complex, high-stakes domains [25]. In related work, we have quantified the impact of automated agents on online discourse health, providing a pedagogical framework for intervening to reduce toxicity [6; 12; 27]. These are blueprints for the future of personalized education: systems that not only possess knowledge but understand the context required to transfer that knowledge to a human partner.

Information Accessibility and Cognitive Health. I am committed to a broad definition of accessibility that encompasses not just physical access, but cognitive and informational equity. Current models often present barriers to users through overly dense text or hallucinatory advice. My work on "speechworthy" language models addresses the cognitive load of audio interfaces, using radio-industry best practices to align output with the natural rhythms of oral communication [16]. Our approach allows information remains accessible to those who rely on non-visual modalities or have difficulty processing dense text.

I am equally committed to community-grounded research to address bias. We pioneered methods to involve threatened communities directly in the creation of bias benchmarks, finding that instruction-tuned models cannot be trusted to self-police without human oversight [19; 20]. Recently, I have extended this to public health. Through the **CHAT-HPV** grant with USC's Keck School of Medicine we are developing conversational agents that bridge the

health literacy gap for underserved populations. We are also rigorously evaluating the safety of these interventions for vulnerable users; my recent work at the NeurIPS GenAI4Health workshop explores intervention opportunities for suicide prevention, ensuring that automated assistants support rather than endanger users in mental health crises [49]. This portfolio aligns directly with AIM's mission to deploy safe, accessible intelligence that serves the full spectrum of human neurodiversity and health needs.

Cognitive Alignment and The Neuroscience of Learning. My long-term research agenda seeks to bridge the gap between computational processing and biological learning. While text is our current primary medium for instruction, it is often a lossy compression of thought. Our recent NeurIPS work, "Language Models Can Predict Their Own Behavior," [3] demonstrates that hidden states in language models can *anticipate* the need to reflect, the ability to comply with instruction, and the confidence of their output, suggesting a form of "thinking" behavior that parallels human cognition. At UMD, I plan to use AIM's interdisciplinary strengths in neuroscience to explore whether these internal model representations can map to human cognitive states. By understanding the "shared geometry" of model and human representation of concepts, we can move beyond text-based instruction toward direct forms of knowledge acquisition. This high-risk trajectory aims to develop non-invasive techniques that align human and machine internal states, potentially revolutionizing how we approach education and skill transfer in the post-digital age.

Institutional Synergy and Leadership. The Artificial Intelligence Interdisciplinary Institute at Maryland (AIM) is the ideal environment for this work, largely because my research is already deeply intertwined with its faculty. I intend to hit the ground running by accelerating my active collaborations, such as the CHIRON and ALLAN programs, which I co-led with Jordan Boyd-Graber to explore agentic deception and negotiation. Furthermore, my current work on the IARPA HIATUS program with Hal Daumé III and Marine Carpuat provides a ready-made foundation for high-impact inquiry into authorship privacy and attribution. Beyond these existing ties, I see immense potential for new synergies with colleagues like Philip Resnik in computational sociolinguistics and Mohit Iyyer in discourse modeling, further cementing AIM's leadership in human-centric NLP.

To formalize these efforts, I plan to re-establish my Center for Useful Techniques Enhancing Language Applications Based on Natural And Meaningful Evidence (CUTELABNAME) at UMD. This center will serve as a nexus for the "human-centric" engineering described above, taking advantage of UMD's proximity to Washington D.C. to influence policy on bias and safety while securing the large-scale government and philanthropic funding required to sustain such an ambitious agenda. I am eager to bring my experience in center-scale leadership and my passion for interdisciplinary inquiry to UMD, helping AIM define the next generation of communicative intelligence.

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] ASHOK, D., AND **MAY, J.** Language models can predict their own behavior. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems* (2025).

[4] 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.

[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] CHIANG, D. Hope and fear for discriminative training of statistical translation models. *Journal of Machine Learning Research 13*, 40 (2012), 1159–1187.

[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] CHO, H. J., SHARMA, K., JEDEMA, N. P., RIBEIRO, L. F. R., **MAY, J.**, AND MOSCHITTI, A. Tuning-free personalized alignment via trial-error-explain in-context learning. In *Findings of the Association for Computational Linguistics: NAACL 2025* (Albuquerque, New Mexico, Apr. 2025), L. Chiruzzo, A. Ritter, and L. Wang, Eds., Association for Computational Linguistics, pp. 5864–5885.

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

[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., 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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. **Best Demo Award**.

[30] 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.

[31] 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.

[32] 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.

[33] Huang, T., Lee, D. H., Sweeney, J., Shi, J., Steliotes, E., Lange, M., **May, J.**, and Chen, M. Foodpuzzle: Toward developing large language model agents as autonomous flavor scientists. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2* (New York, NY, USA, 2025), KDD '25, Association for Computing Machinery, p. 5493–5504.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] Liu, S., and **May, J.** Style transfer with multi-iteration preference optimization, 2024.

[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] 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.

[41] 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).

[42] **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.

[43] 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.

[44] 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.

[45] Och, F. J. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics* (Sapporo, Japan, July 2003), Association for Computational Linguistics, pp. 160–167.

[46] 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.

[47] 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).

[48] 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.

[49] Ranjit, J., Cho, H. J., Smerdon, C. J., Nam, Y., Phung, M., **May, J.**, Blosnich, J. R., and Swayamdipta, S. Supporting the design of intervention opportunities for suicide prevention with language model assistants, 2025.

[50] 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**.

[51] 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.

[52] 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.

[53] 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**.

[54] 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.

[55] 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.

[56] 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.

[57] 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.