

On Foundation Models for Temporal Point Processes to Accelerate Scientific Discovery

David Berghaus^{1,2}, Patrick Seifner^{1,3}, Kostadin Cvejovski^{1,4},
& Ramsés J. Sánchez^{1,2,3}

¹Lamarr Institute, ²Fraunhofer IAIS, ³University of Bonn, ⁴JetBrains Research

Abstract

Many scientific fields, from medicine to seismology, rely on analyzing sequences of events over time to understand complex systems. Traditionally, machine learning models must be built and trained from scratch for each new dataset, which is a slow and costly process. We introduce a new approach: a single, powerful model that learns the underlying patterns of event data *in-context*. We trained this "foundation model" on millions of simulated event sequences, teaching it a general-purpose understanding of how events can unfold. As a result, our model can analyze new scientific data instantly, without retraining, simply by looking at a few examples from the dataset. It can also be quickly fine-tuned for even higher accuracy. This approach makes sophisticated event analysis more accessible and accelerates the pace of scientific discovery. Our pretrained model, repository and tutorials **will soon be available online**¹.

Keywords: Temporal Point Processes, Foundation Models, In-Context Learning

1 Introduction and Related Work

The pace of scientific discovery is being dramatically accelerated by artificial intelligence, with breakthrough applications emerging across numerous fields. A landmark example is AlphaFold [Jumper et al., 2021], which has revolutionized structural biology by accurately predicting protein structures.

This success is part of a broader trend where AI is becoming an indispensable tool for researchers, enabling them to tackle previously intractable problems in fields like weather-forecasting [Price et al., 2025], material science [Merchant et al., 2023], medicine [Poplin et al., 2018, Enguehard et al., 2020, Berger et al., 2024a,b, 2025] and mathematics [Novikov et al., 2025, Berghaus, 2025].

A common challenge in these and many other scientific domains is the need to model *sequences of discrete events that occur at irregular intervals in continuous time* [Cvejovski et al., 2020, 2021, Ojeda et al., 2021, Seifner and Sanchez, 2023, Khanna et al., 2025]. Temporal point processes (TPPs) provide a powerful and principled mathematical framework for this task [Daley and Vere-Jones, 2003]. By capturing the underlying dynamics of event streams — not just *when* events happen, but also *what* they are — TPPs enable us to forecast future events, understand complex dependencies, and gain deeper insights into the phenomena being studied.

¹<https://fim4science.github.io/OpenFIM/intro.html>

For this reason, various works have modeled TPPs using deep learning techniques [Du et al., 2016, Mei and Eisner, 2017, Kim et al., 2019, Panos et al., 2023].

However, a fundamental limitation of these approaches is their *lack of transferability*: a new TPP model must be trained from scratch for each new dataset, a slow and costly process. To overcome this, a new paradigm has emerged: pretraining a single, powerful model on vast amounts of synthetic data that can then perform inference *in-context* (or *zero-shot*) on new, unseen problems. Within this paradigm, two variants have recently emerged: Prior-fitted Networks (PFNs) and Foundation Inference Models (FIMs). PFNs train networks to approximate *predictive posterior distributions* in a sequence-to-sequence fashion, often implemented with transformer architectures [Müller et al., 2022, Hollmann et al., 2022, Müller et al., 2025]. FIMs, by contrast, focus on directly *estimating the infinitesimal operators* of stochastic processes, thereby retaining a degree of interpretability [Berghaus et al., 2025a, Seifner et al., 2025a,b, Mauel et al., 2025, Hinz et al., 2025].

In this work, we demonstrate the first Foundation Inference Model for Marked Temporal Point Processes, the complete details of which will be presented elsewhere [Berghaus et al., 2025b]. Our model learns to directly estimate the conditional intensity function — the fundamental operator governing TPP dynamics — from a small context of observed event sequences. This allows practitioners to obtain accurate, *zero-shot* characterizations of event dynamics, accelerating the pace of scientific discovery.

2 In-Context TPP Inference

We propose a novel in-context learning method for MTPP intensity inference. First, we generate a large set of marked event sequences from a broad distribution over MTPPs. This yields training data for a neural network recognition model, which is trained to estimate the underlying ground-truth intensity function via maximum likelihood. Such a pretrained inference model can then be applied directly to real-world problems (zero-shot) or rapidly finetuned to target data.

2.1 Synthetic Dataset Generation

To design a synthetic dataset, we define a *prior distribution over Hawkes processes*. An instance is characterized by a conditional intensity function for each mark κ :

$$\lambda(t, \kappa \mid \mathcal{H}_t) = \max \left(0, \mu_\kappa(t) + \sum_{(t', \kappa') \in \mathcal{H}_t} z_{\kappa\kappa'} \gamma_{\kappa\kappa'}(t - t') \right). \quad (1)$$

We generate a diverse dataset by sampling the functional forms and parameters for the base intensities μ_κ and interaction kernels $\gamma_{\kappa\kappa'}$ from broad distributions (e.g., constant, sinusoidal, exponential decay). The factors $z_{\kappa\kappa'} \in \{-1, 0, 1\}$ are sampled to create excitatory, inhibitory, or neutral interactions with varying levels of sparsity. From each sampled intensity function, we simulate a collection of event sequences using Ogata’s thinning algorithm [Ogata, 1981].

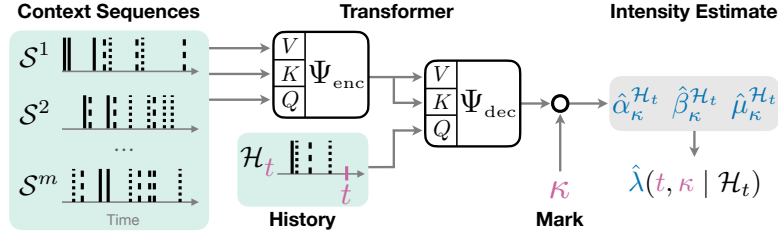


Figure 1: Schematic of FIM-PP [Berghaus et al., 2025b]. A *context* of event sequences is encoded by a transformer encoder. The result is used as memory by a transformer decoder, which takes a specific event *history* $\mathcal{H}_t$ as a query. The final embedding, combined with a target *mark* κ , is projected to parameters that define the *conditional intensity function* $\hat{\lambda}(t, \kappa | \mathcal{H}_t)$.

2.2 Foundation Inference Model Architecture

FIM-PP is a pretrained deep neural network that processes a *context* of event sequences $\mathcal{C} = \{\mathcal{S}^j\}_{j=1}^m$ to estimate the conditional intensity function $\hat{\lambda}$ that describes the observed dynamics. As shown in Figure 1, the model uses an architecture that mimics the original transformer [Vaswani et al., 2017].

First, each sequence in the context set $\mathcal{C}$ is embedded and processed by a transformer encoder to produce a summary vector. These summary vectors are then passed through another transformer encoder to share information across sequences, yielding a final context representation. To estimate the intensity at time t , an event history $\mathcal{H}_t$ is embedded and used as the query for a transformer decoder, which attends to the context representation. The output of the decoder is a context-aware history embedding $\mathbf{h}_t^{\text{hist}}$.

This embedding is then used to parameterize a flexible intensity function for each mark κ' :

$$\hat{\lambda}(t, \kappa' | \mathcal{H}_t) = \hat{\mu}_{\kappa'}^{\mathcal{H}_t} + (\hat{\alpha}_{\kappa'}^{\mathcal{H}_t} - \hat{\mu}_{\kappa'}^{\mathcal{H}_t}) \exp\left(-\hat{\beta}_{\kappa'}^{\mathcal{H}_t}(t - t_{n_{\text{hist}}})\right), \quad (2)$$

where $t_{n_{\text{hist}}}$ is the time of the last event in the history, and the non-negative parameters $(\hat{\mu}_{\kappa'}^{\mathcal{H}_t}, \hat{\alpha}_{\kappa'}^{\mathcal{H}_t}, \hat{\beta}_{\kappa'}^{\mathcal{H}_t})$ are predicted by small MLPs from $\mathbf{h}_t^{\text{hist}}$. The model is trained by minimizing the negative log-likelihood (NLL) of a target sequence, given the remaining sequences as context.

3 Demonstration of Capabilities: Preliminary Results

To demonstrate the effectiveness of our foundation model, we visualize its zero-shot performance on several synthetic temporal point processes. The following examples showcase the capabilities of the pretrained model, FIM-PP, applied directly to new data without any dataset-specific training.

Figure 2 visually confirms the model’s core ability to infer the conditional intensity of different underlying processes **in a zero-shot manner**. The plots show that FIM-PP

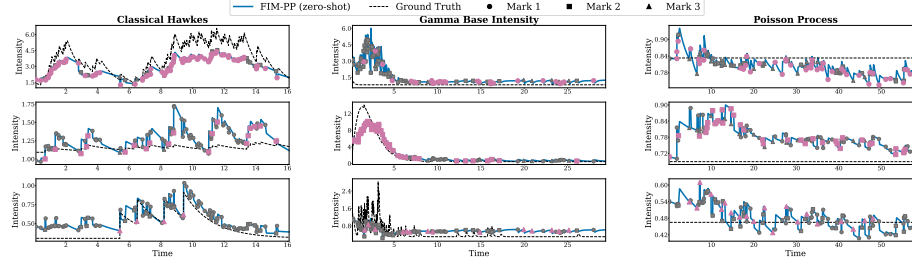


Figure 2: Demonstration of FIM-PP’s *zero-shot* inference capabilities on examples from three synthetic datasets: a synthetic Poisson process, a process with a constant base intensity, and one with a gamma-distributed base intensity. The zero-shot model effectively captures the underlying dynamics by leveraging its pretrained prior. The ground truth intensity is shown alongside the model’s estimate, illustrating a close match even without dataset-specific training.

produces remarkably accurate estimates that closely track the ground-truth dynamics for a variety of processes. This powerful generalization capability stems directly from its pretraining on a diverse corpus of synthetic Hawkes processes.

After downloading the model and providing it a few example sequences as context, a user can instantly obtain this intensity function for any given event history. This function can then be used for various downstream tasks, such as forecasting the time and type of the next event, simulating future event trajectories, or analyzing the excitatory and inhibitory influences between different event types. Evaluations of FIM-PP on real-world datasets can be found in [Berghaus et al., 2025b].

4 Conclusion

In this work, we introduced FIM-PP, the first Foundation Inference Model capable of inferring MTPPs from real-world data. Our experiments demonstrated that a single FIM-PP, pretrained on synthetic Hawkes-process data only, can match the predictive performance of specialized MTPP methods in zero-shot mode. Finetuning on target data further boosts performance, enabling FIM-PP to outperform competing methods across the majority of evaluated tasks [Berghaus et al. 2025b].

Collaboration: We are very interested in collaboration. We would be especially interested in solving real world problems with FIM-PP and encourage practitioners from any domain to reach out to us.

Acknowledgments

This research has been funded by the Federal Ministry of Education and Research of Germany and the state of North-Rhine Westphalia as part of the Lamarr Institute for Machine Learning and Artificial Intelligence.

References

- A. Berger, A. H. Bashir, D. Berghaus, Mowmita, N. Afsan, L. Grigull, L. Fendrich, H. Högl, G. Ernst, R. Schmidt, D. Bascom, T. A. Lagones, T. Deußner, T. Bell, M. Lübbering, and R. Sifa. Optimizing rare disease patient matching with large language models. In *2024 IEEE International Conference on Big Data (BigData)*, pages 1–5, 2024a. doi: 10.1109/BigData62323.2024.10910113.
- A. Berger, D. Berghaus, A. H. Bashir, L. Grigull, L. Fendrich, T. A. Lagones, H. Högl, G. Ernst, R. Schmidt, D. Bascom, T. Deußner, T. Bell, M. Lübbering, and R. Sifa. Advancing personalized medicine: A scalable llm-based recommender system for patient matching. In *2024 IEEE International Conference on Big Data (BigData)*, pages 5876–5883, 2024b. doi: 10.1109/BigData62323.2024.10825567.
- A. Berger, S. Khanna, D. Berghaus, and R. Sifa. Reasoning llms in the medical domain: A literature survey, 2025. URL <https://arxiv.org/abs/2508.19097>.
- D. Berghaus. Not constructing ramsey graphs using deep reinforcement learning. In *I can't believe it's not better (ICLR)*, 2025.
- D. Berghaus, K. Cvejowski, P. Seifner, C. Ojeda, and R. J. Sanchez. Foundation inference models for markov jump processes. *Proceedings of the 38th International Conference on Neural Information Processing Systems*, 2025a.
- D. Berghaus, P. Seifner, K. Cvejowski, C. Ojeda, and R. J. Sánchez. In-context learning of temporal point processes with foundation inference models, 2025b. URL <https://arxiv.org/abs/2509.24762>.
- K. Cvejowski, R. J. Sánchez, B. Georgiev, C. Bauckhage, and C. Ojeda. Recurrent point review models. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.
- K. Cvejowski, R. J. Sanchez, C. Bauckhage, and C. Ojeda. Dynamic review-based recommenders. In *International Data Science Conference*, 2021. arXiv:2110.14747.
- D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, second edition, 2003.
- N. Du, H. Dai, R. Trivedi, U. Upadhyay, M. Gomez-Rodriguez, and L. Song. Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1555–1564, 2016.
- J. Enguehard, D. Busbridge, A. Bozson, C. Woodcock, and N. Y. Hammerla. Neural temporal point processes for modelling electronic health records, 2020. URL <https://arxiv.org/abs/2007.13794>.
- M. Hinz, M. Mauel, P. Seifner, D. Berghaus, K. Cvejowski, and R. Sanchez. Towards fast coarse-graining and equation discovery with foundation inference models, 2025.

- N. Hollmann, S. Müller, K. Eggensperger, and F. Hutter. TabPFN: A transformer that solves small tabular classification problems in a second. *arXiv preprint arXiv:2207.01848*, 2022.
- J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2.
- S. Khanna, A. Berger, D. Berghaus, T. Deusser, L. Sparrenberg, and R. Sifa. Towards unified multimodal financial forecasting: Integrating sentiment embeddings and market indicators via cross-modal attention, 2025. URL <https://arxiv.org/abs/2508.13327>.
- H. Kim, A. Mnih, J. Schwarz, M. Garnelo, A. Eslami, D. Rosenbaum, O. Vinyals, and Y. W. Teh. Attentive neural processes. *arXiv preprint arXiv:1901.05761*, 2019.
- M. Mauel, M. Hinz, P. Seifner, D. Berghaus, and R. Sánchez. Towards foundation inference models that learn odes in-context, 2025.
- H. Mei and J. M. Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/6463c88460bd63bbe256e495c63aa40b-Paper.pdf.
- A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, dec 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06735-9. URL <https://doi.org/10.1038/s41586-023-06735-9>.
- S. Müller, N. Hollmann, S. P. Arango, J. Grabocka, and F. Hutter. Transformers can do bayesian inference. In *International Conference on Learning Representations*, 2022. URL <https://openreview.net/forum?id=KSugKcbNf9>.
- S. Müller, A. Reuter, N. Hollmann, D. Rügamer, and F. Hutter. Position: The future of bayesian prediction is prior-fitted. *arXiv preprint arXiv:2505.23947*, 2025.
- A. Novikov, N. Vü, M. Eisenberger, E. Dupont, P.-S. Huang, A. Z. Wagner, S. Shirobokov, B. Kozlovskii, F. J. R. Ruiz, A. Mehrabian, M. P. Kumar, A. See, S. Chaudhuri, G. Holland, A. Davies, S. Nowozin, P. Kohli, and M. Balog. Alphaevolve: A coding agent for scientific and algorithmic discovery, 2025. URL <https://arxiv.org/abs/2506.13131>.
- Y. Ogata. On lewis’ simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.

- C. Ojeda, K. Cvejoski, B. Georgiev, C. Bauckhage, J. Schuecker, and R. J. Sánchez. Learning deep generative models for queuing systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9214–9222, 2021.
- A. Panos, I. Kosmidis, and P. Dellaportas. Scalable marked point processes for exchangeable and non-exchangeable event sequences. In F. Ruiz, J. Dy, and J.-W. van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 236–252. PMLR, 25–27 Apr 2023. URL <https://proceedings.mlr.press/v206/panos23a.html>.
- R. Poplin, P.-C. Chang, D. Alexander, S. Schwartz, T. Colthurst, A. Ku, D. Newburger, J. Dijamco, N. Nguyen, P. T. Afshar, S. S. Gross, L. Dorfman, C. Y. McLean, and M. A. DePristo. A universal snp and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36(10):983–987, oct 2018. ISSN 1546-1696. doi: 10.1038/nbt.4235. URL <https://doi.org/10.1038/nbt.4235>.
- I. Price, A. Sanchez-Gonzalez, F. Alet, T. R. Andersson, A. El-Kadi, D. Masters, T. Ewalds, J. Stott, S. Mohamed, P. Battaglia, R. Lam, and M. Willson. Probabilistic weather forecasting with machine learning. *Nature*, 637(8044):84–90, 2025. doi: 10.1038/s41586-024-08252-9.
- P. Seifner and R. J. Sanchez. Neural Markov jump processes. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202. PMLR, 2023.
- P. Seifner, K. Cvejoski, D. Berghaus, C. Ojeda, and R. J. Sanchez. Foundation inference models for stochastic differential equations: A transformer-based approach for zero-shot function estimation, 2025a. URL <https://arxiv.org/abs/2502.19049>.
- P. Seifner, K. Cvejoski, A. Körner, and R. J. Sanchez. Zero-shot imputation with foundation inference models for dynamical systems. In *The Thirteenth International Conference on Learning Representations*, 2025b. URL <https://openreview.net/forum?id=NPSZ7V1CCY>.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.