# In-Context Learning of Temporal Point Processes with Foundation Inference Models

David Berghaus[1,2] Patrick Seifner[1,3] Kostadin Cvejoski[4] César Ojeda[5] Ramsés J. Sánchez[1,2,3]

[1]Lamarr Institute [2]Fraunhofer IAIS [3]University of Bonn [4]JetBrains Research [5]University of Potsdam

*Contact:* david.berghaus@iais.fraunhofer.de *Project:* fim4science.github.io/OpenFIM/intro.html

## Problem & Approach

**Marked temporal point processes (MTPPs)** model irregular event streams in which each event has both a *time* and a *type*. Their central object is the conditional intensity $\lambda(t, k \mid \mathcal{H}_t)$, which describes how likely an event of mark $k$ is to occur at time $t$ given the past history $\mathcal{H}_t = \{(t_i, k_i) \mid t_i < t\}$.

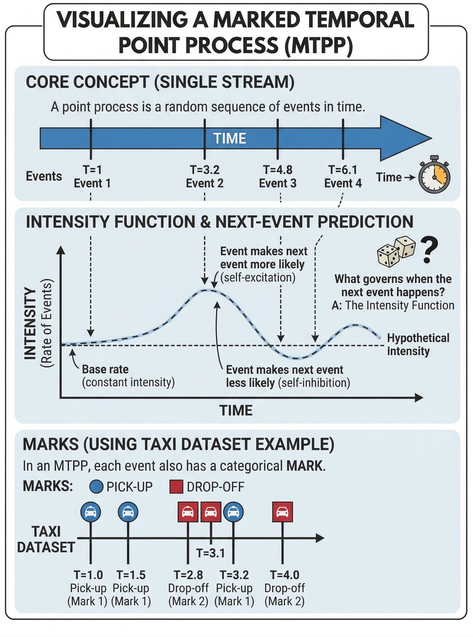
**Current limitation.** Most neural temporal point-process models are trained *dataset by dataset*: for every new system, they must be retrained from scratch. This prevents **zero-shot transfer** across dynamical regimes and limits practical reuse.

**Our idea. FIM-PP**is a *pretrained recognition model* for point processes. It is trained on a broad synthetic prior of marked point processes and learns to infer conditional intensities *in context* from a set of related event sequences.

**Why this matters.** A single pretrained model can be applied directly to new event data, retains access to an explicit and interpretable intensity estimate, and can be improved further with rapid fine-tuning.

### Takeaway

Pretrain once on synthetic event dynamics; then infer new marked point-process behavior zero-shot or adapt within minutes.




## Foundation Inference Models

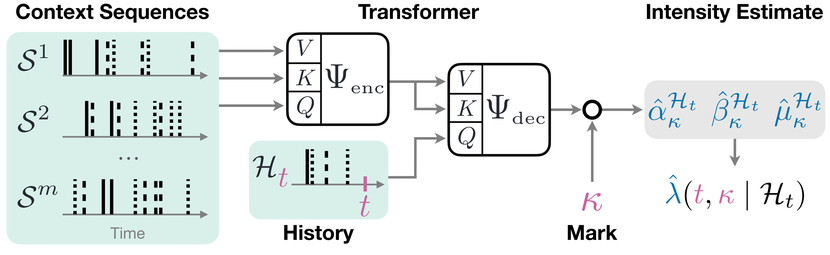
**Part A: Synthetic training data generation**

We sample processes from a broad prior over conditional intensities

$$\lambda(t, k \mid \mathcal{H}_t) = \max\left(0, \mu_k(t) + \sum_{(t',k')\in\mathcal{H}_t} z_{kk'}\gamma_{kk'}(t-t')\right),$$

where $\mu_k(t)$ is chosen as either a **constant**, a **positive sinusoid**, or a **Gamma-shaped initial-rate function**, and $\gamma_{kk'}(t)$ is chosen as either **zero**, an **exponential Hawkes kernel**, or a **shifted Rayleigh kernel**. Interaction signs $z_{kk'} \in \{-1, 0, 1\}$ cover *excitatory*, *inhibitory*, and *neutral* relations.

**Part B: Amortized inference model**




*Context is encoded first, target-history queries then attend to it, and a mark-conditioned head predicts intensity parameters.*

- Context sequences are embedded from event times, marks, and inter-event times.
- The target history queries the learned context representation.
- The final mark-conditioned output predicts an interpretable intensity function.

**Part C: Intensity head**

The model outputs a mark-wise conditional intensity, which can directly show self-excitation, inhibition, and time-varying event likelihood.

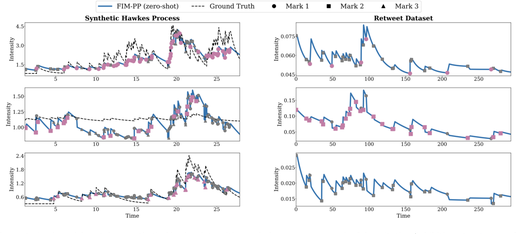



Figure: The pretrained model matches synthetic Hawkes intensities and also recovers meaningful excitation/inhibition structure on **Retweet**.

**Part D: Training and fine-tuning**

Pretraining uses synthetic sequences and next-event likelihood. At application time, the model is used directly in **FIM-PP** (zero-shot)mode or adapted on the target dataset in **FIM-PP** (fine-tuned)mode, which typically takes only a few minutes and less than 11 GB of GPU memory.

## Long-Horizon Prediction: $N = 20$

| | Taxi | | StackOverflow | | Amazon | | Retweet | |
|---|---|---|---|---|---|---|---|---|
| Method | OTD | sMAPE | OTD | sMAPE | OTD | sMAPE | OTD | sMAPE |
| HYPRO | 21.60 | 93.8 | 42.40 | 111.0 | 38.6 | 82.5 | 61.03 | 106.11 |
| Dual-TPP | 24.48 | 95.2 | 41.75 | 117.58 | 42.6 | 86.5 | 61.10 | 106.90 |
| A-NHP | 24.76 | 97.4 | 42.59 | 108.54 | 39.5 | 84.3 | 60.63 | 107.23 |
| NHP | 25.11 | 96.5 | 43.79 | 116.95 | 42.6 | 92.1 | 60.95 | 107.08 |
| IFTPP | 24.05 | 95.7 | 46.28 | 115.12 | 43.8 | 90.9 | 61.72 | 106.71 |
| TCDDM | 22.15 | 90.6 | 42.13 | 107.66 | 42.2 | 83.8 | 60.50 | 106.05 |
| CDiff | 21.01 | 88.0 | 41.25 | 106.18 | 37.7 | 82.0 | 60.66 | 106.18 |
| **FIM-PP** (zero-shot) | 23.15 | **76.8** | 49.26 | 96.36 | 46.2 | 128.6 | 60.24 | 99.07 |
| **FIM-PP** (fine-tuned) | **17.91** | 76.8 | **39.80** | **88.25** | **37.2** | **81.2** | **59.44** | **87.59** |

## Next-Event Prediction

| | Taxi | | | Taobao | | |
|---|---|---|---|---|---|---|
| Method | $\text{RMSE}_{\Delta t}$ | Acc | $\text{sMAPE}_{\Delta t}$ | $\text{RMSE}_{\Delta t}$ | Acc | $\text{sMAPE}_{\Delta t}$ |
| A-NHP | **0.32** | **0.91** | **85.13** | 0.53 | 0.47 | 129.13 |
| Dual-TPP | 0.34 | **0.91** | 89.12 | 0.53 | 0.47 | 131.43 |
| NHP | 0.34 | **0.91** | 90.63 | 0.53 | 0.46 | 133.69 |
| IFTPP | 0.38 | 0.90 | 90.03 | 0.53 | 0.45 | 126.01 |
| CDiff | 0.34 | **0.91** | 87.12 | 0.52 | 0.48 | 127.12 |
| **FIM-PP** (zero-shot) | 0.34 | 0.47 | 98.03 | **0.13** | 0.52 | **112.89** |
| **FIM-PP** (fine-tuned) | 0.45 | **0.91** | 85.91 | 0.16 | **0.60** | 115.98 |