ML4PM, Copehnagen, 14 October 2024

Predictive Process Monitoring
the story so far and trends for the future

Chiara Ghidini, Free University of Bozen - Bolzano
Massimiliano Ronzani, Fondazione Bruno Kessler - Trento

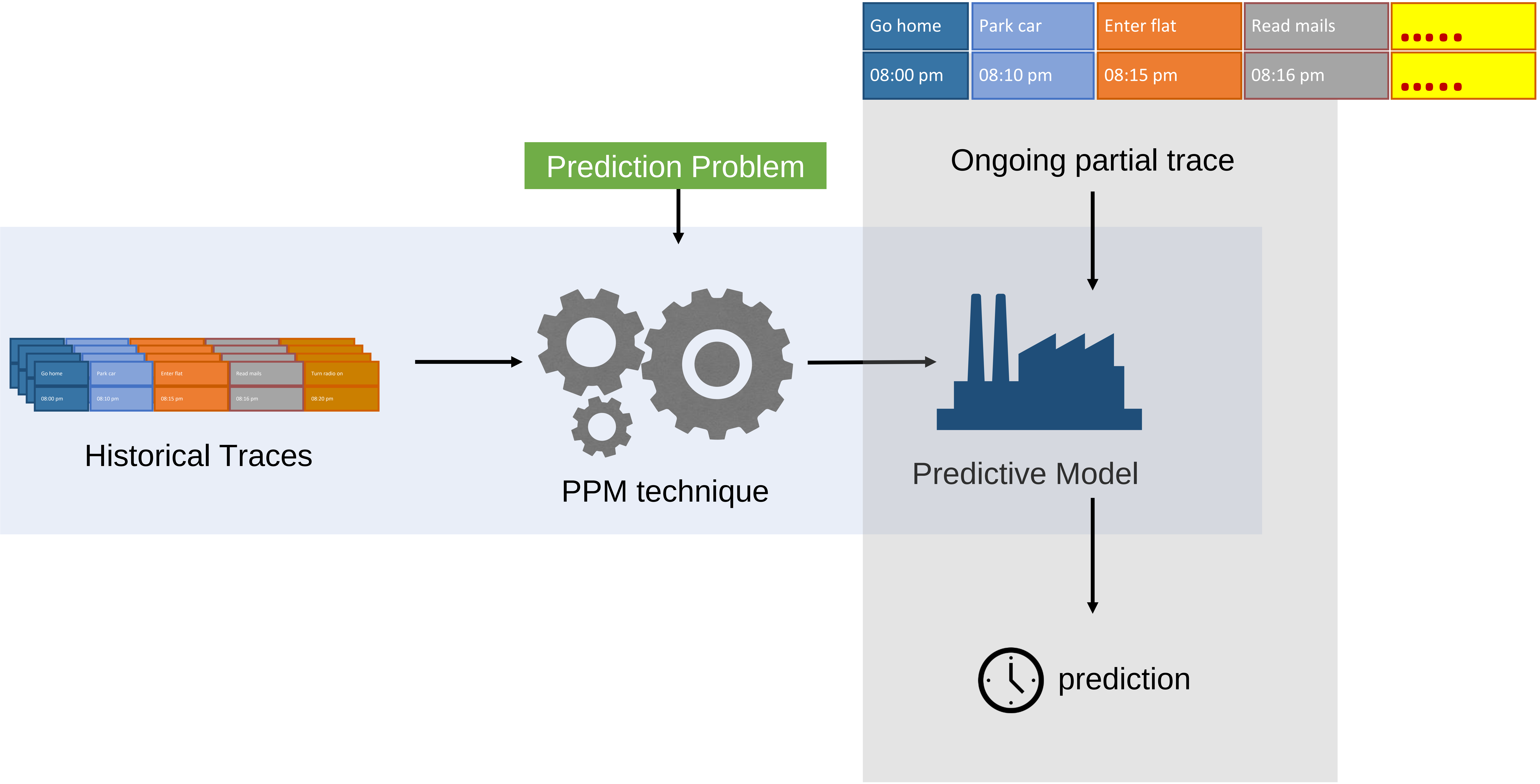
Outline

- The story so far:
  - **Predictive** Process Monitoring so far, or Process Mining and **Supervised ML**
- Trends for the future:
  - Max: **how trends in AI can affect Predictive Process Monitoring**
  - Chiara: **what I would like to see**

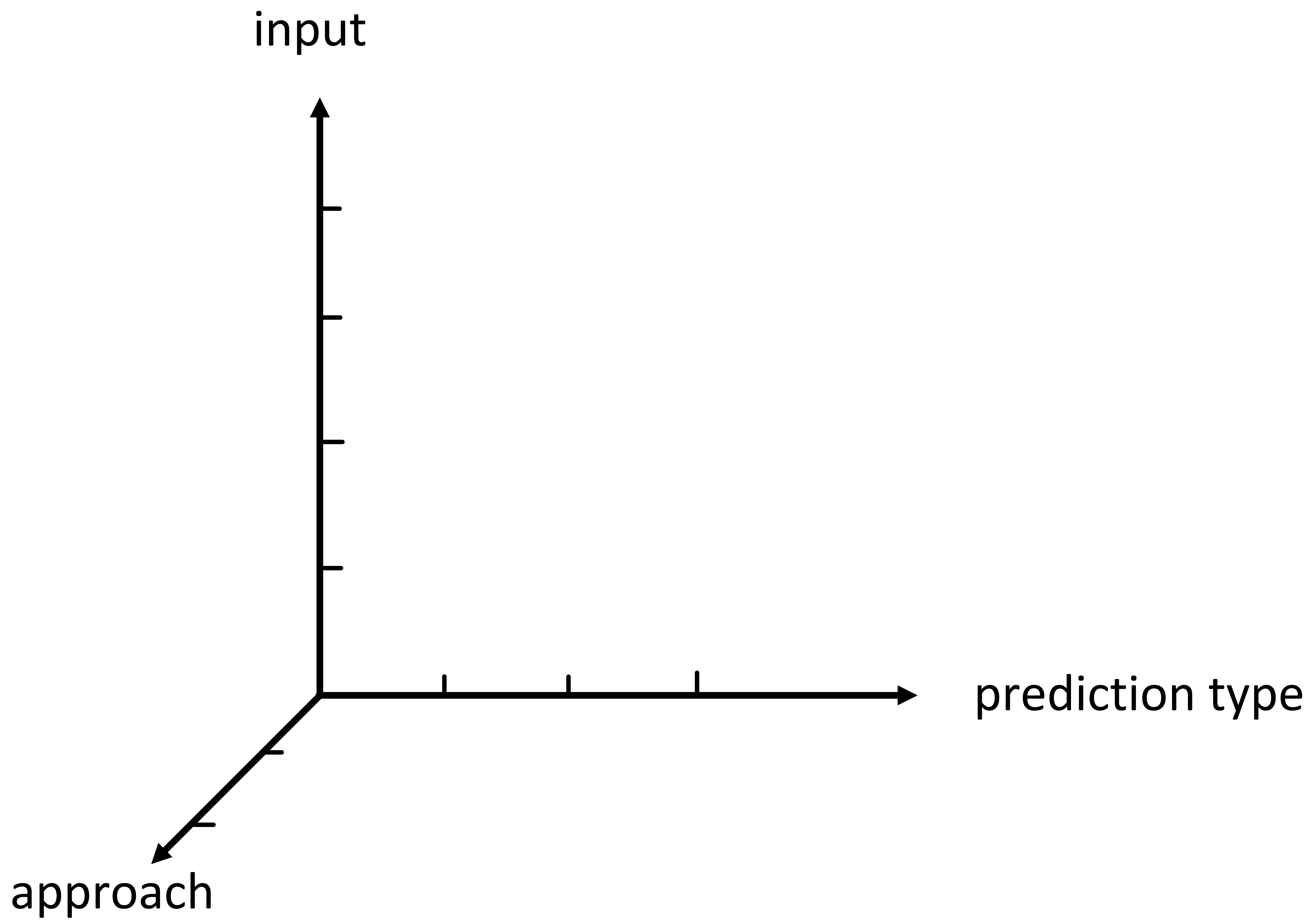
Predictive Process Monitoring

An introduction

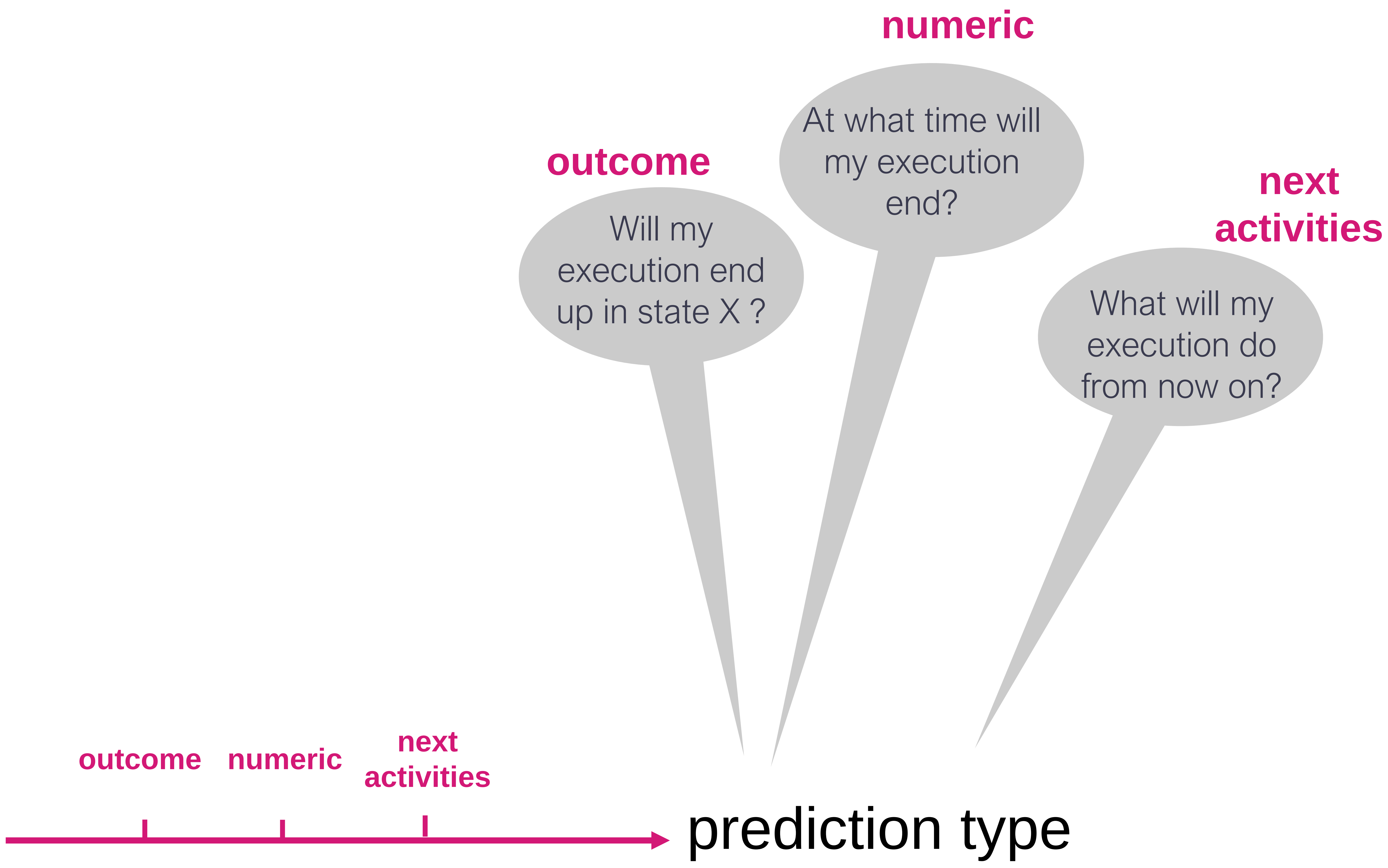
Predictive Process Monitoring (PPM)



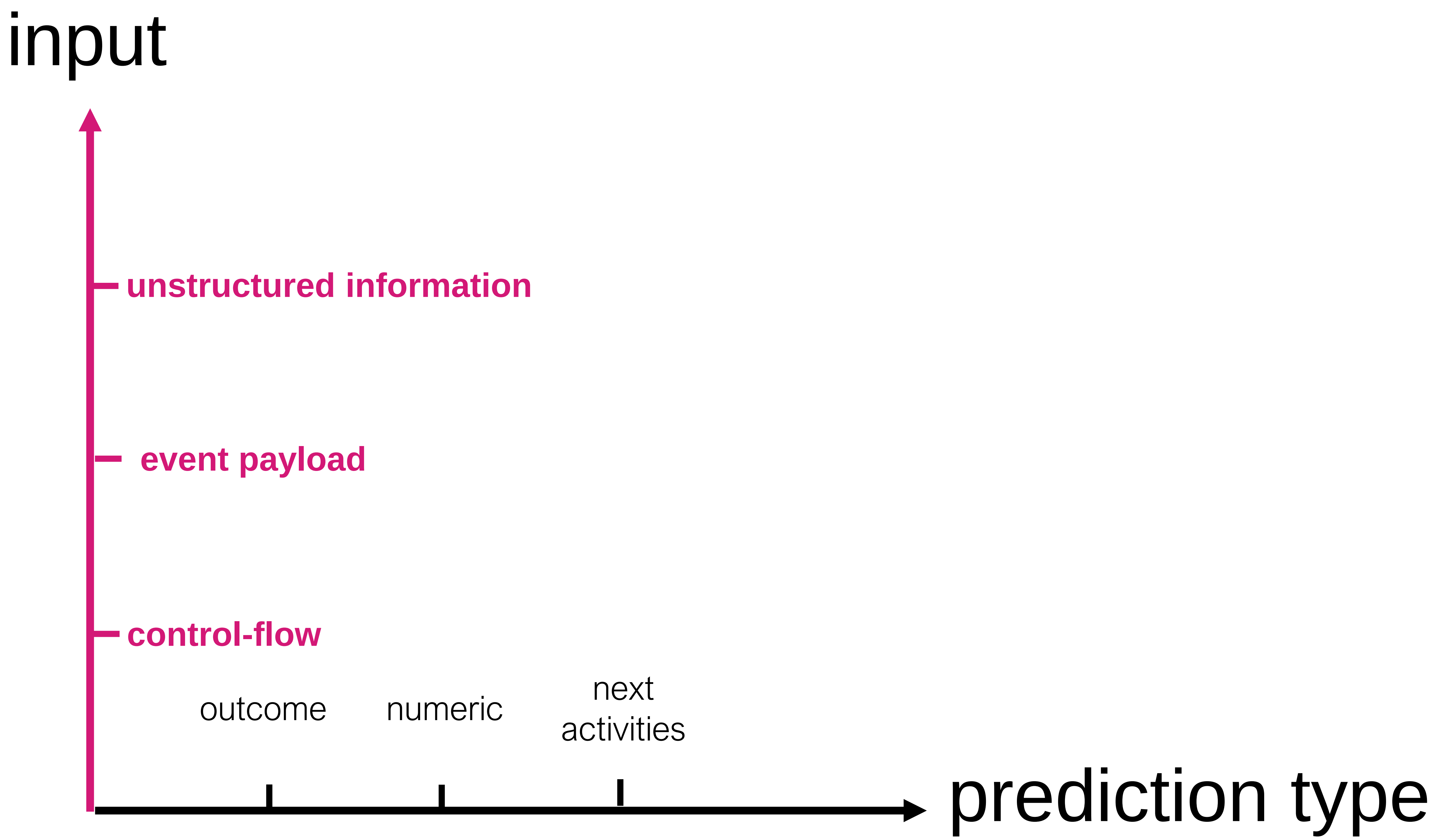
Three dimensions



Dimension 1: what to predict



Dimension 2: which information to use

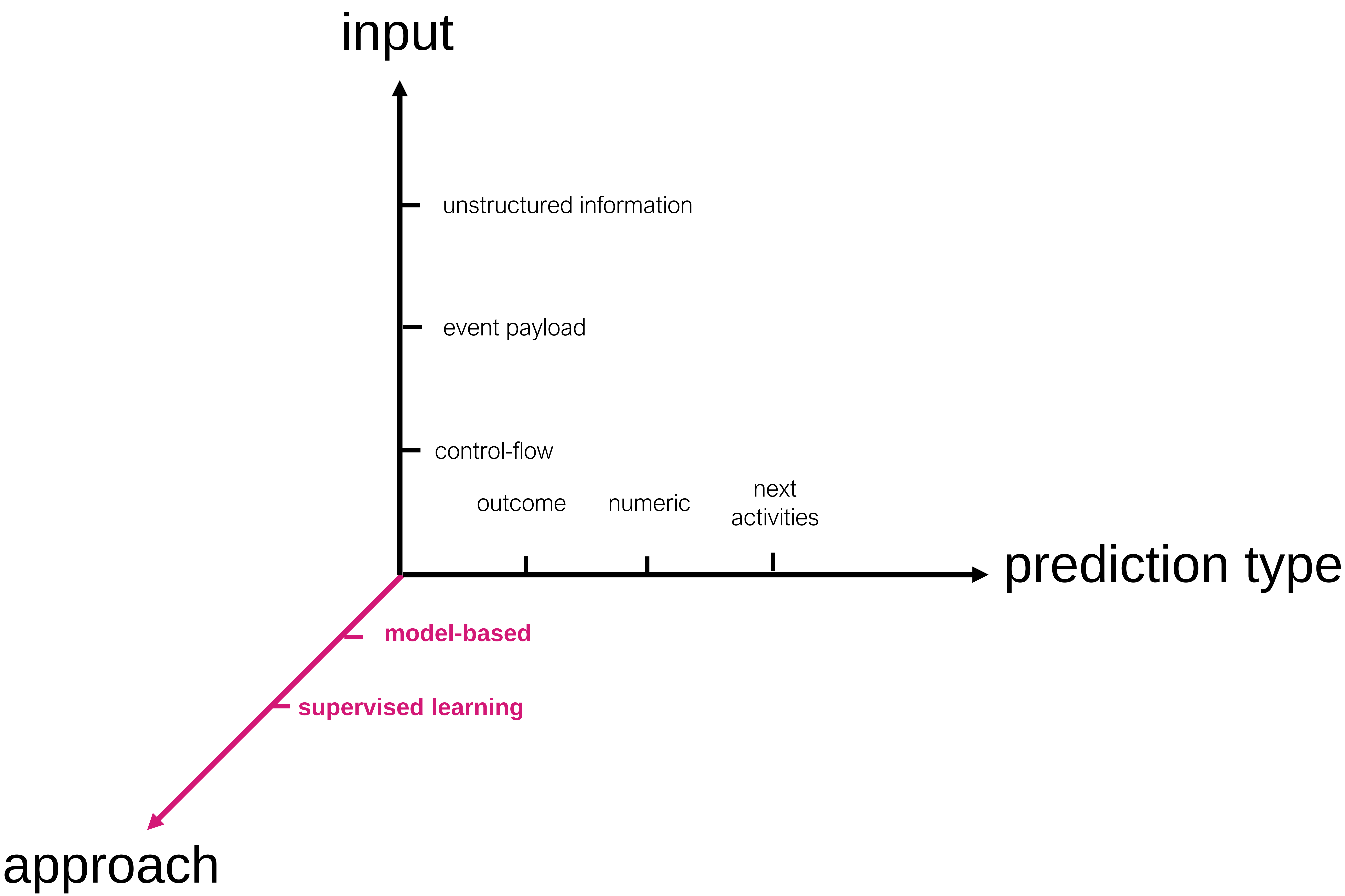


Different types of input information

Patient’s history (trace)

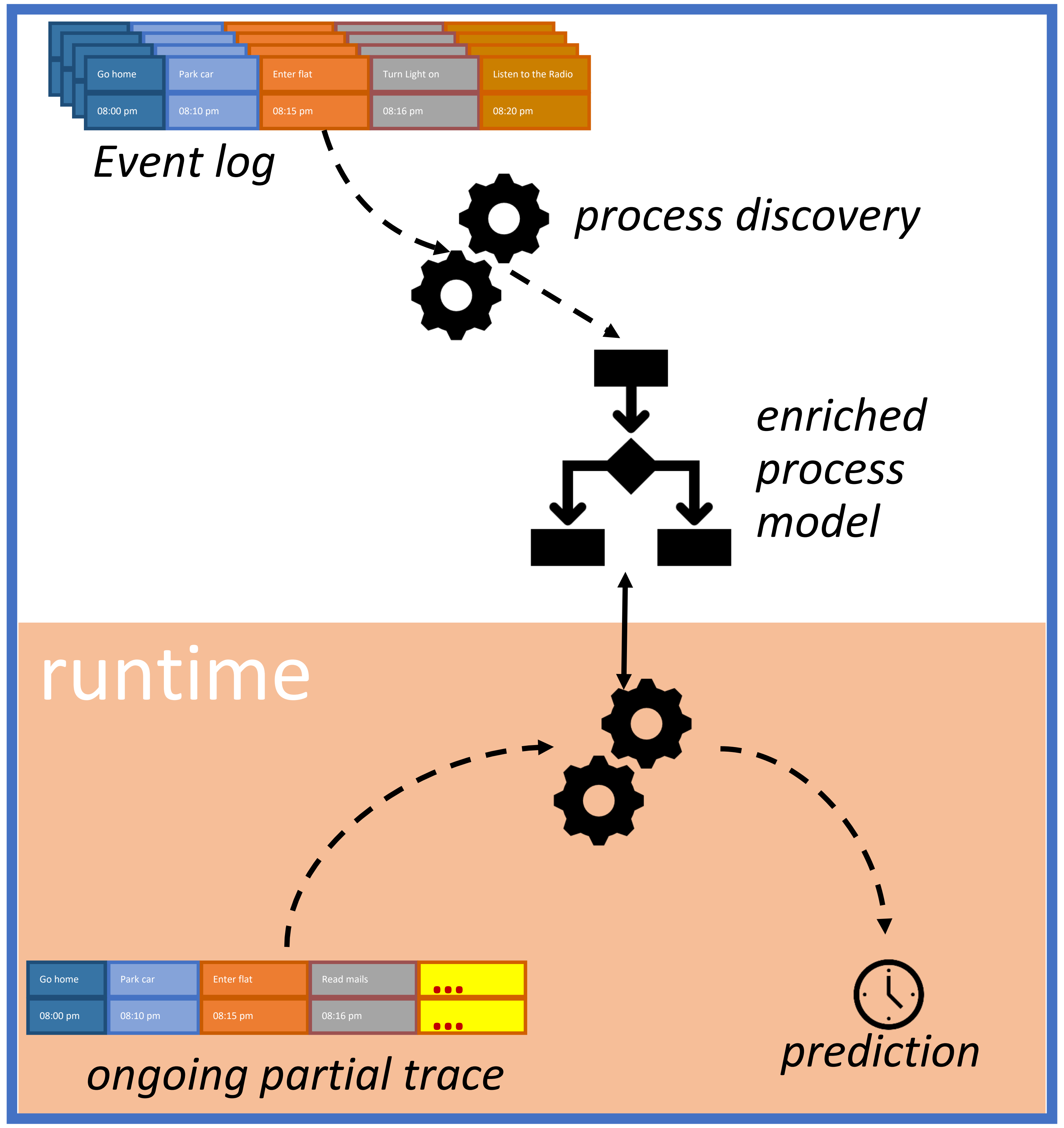
Control flow	Blood test	X-Ray	Diagnosis	Manipulation	Check Visit
Payloads	8/09/2017 Bilirubin: 1.9 mg/DL Calcium: 8.0 mg/D	10/09/2107 Result: Spine abnormally curved	15/09/2017 diagnosis:Scoliosis	20/09/2017 Duration: 10 min	30/09/2017 Exit: recovered
Unstructured content			“The patient presents also a light form of lordosis»	“The patient felt some pain during the treatment”	

Dimension 3: which type of technique



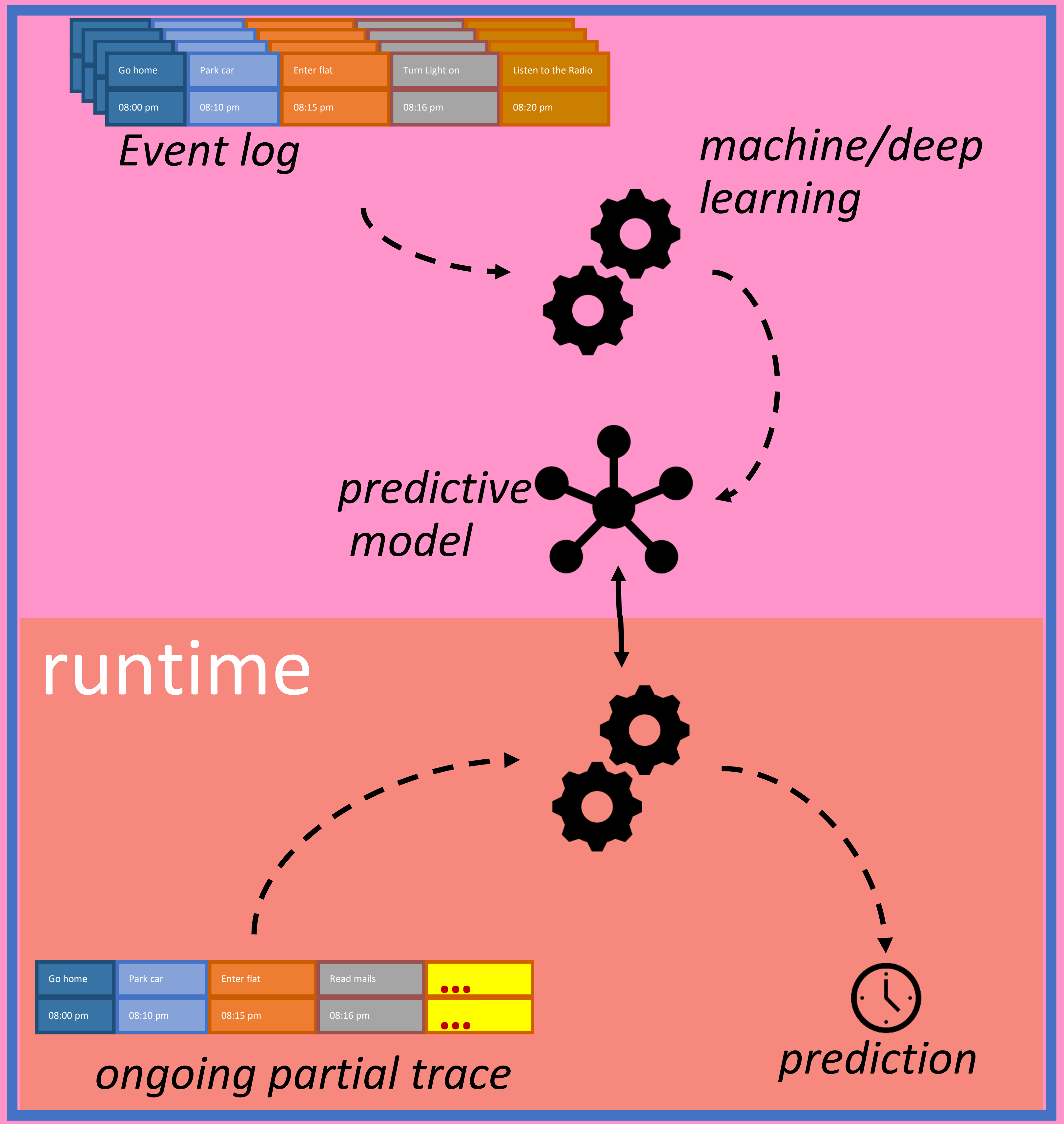
Type of technique

Model-based



In literature, mainly …

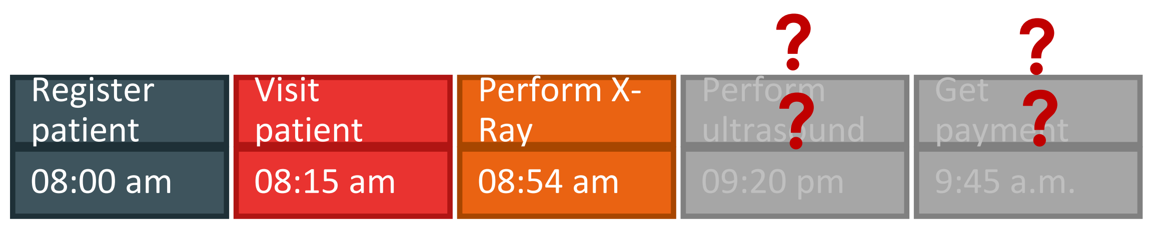
Supervised Learning



The “traditional” pipeline

The “traditional” pipeline

Why Prefixes?



Ongoing partial trace

- The partial ongoing trace is incomplete
- The predictive model needs to learn the **correlations** between **incomplete traces** (trace prefixes) and the **output** that you want to predict
- Formally, it needs to learn a function

$$f(L, \sigma_i^m) = \overline{label_i}$$

So, .. we need to use prefix(es) to train models

Which prefix(es) for which model?

The choice of prefix(es) depends upon:

- the **type of prediction** you are interested in

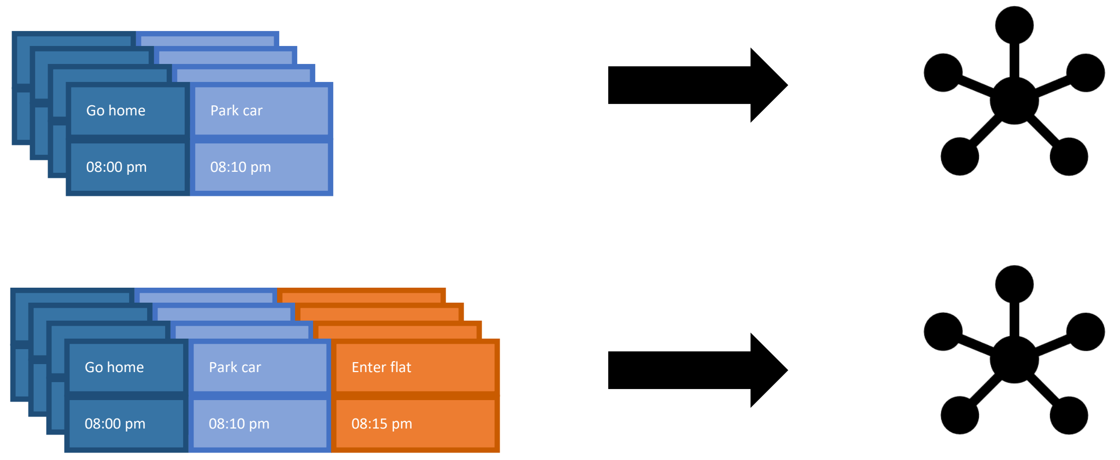
  Outcome prediction vs
  Time prediction of *e* happening in the middle of the process

- **how early** you want to get a prediction

  At the beginning vs late in the process

Which prefix(es) for which model?

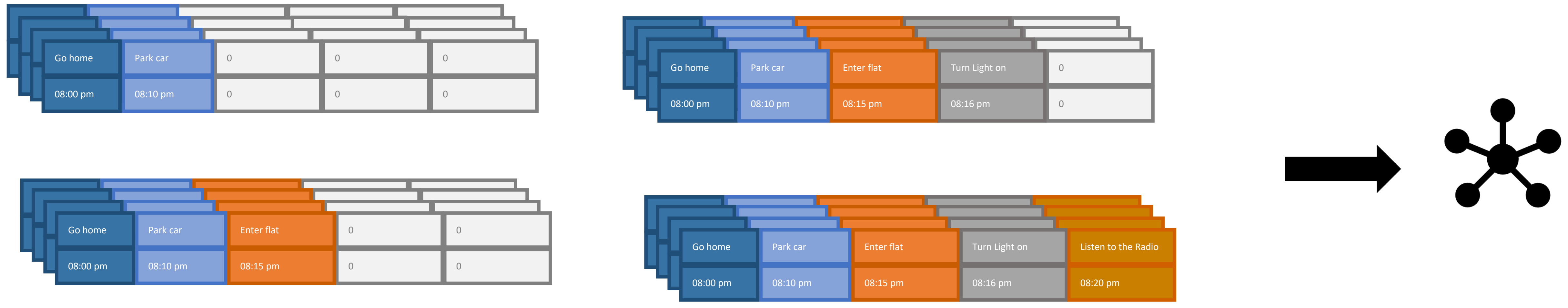
- Predictive model for a **specific prefix length**



Predictive model for prefix length 2

Predictive model for prefix length 3

- Single model **with prefixes of different lengths** altogether
  - using padding if needed -



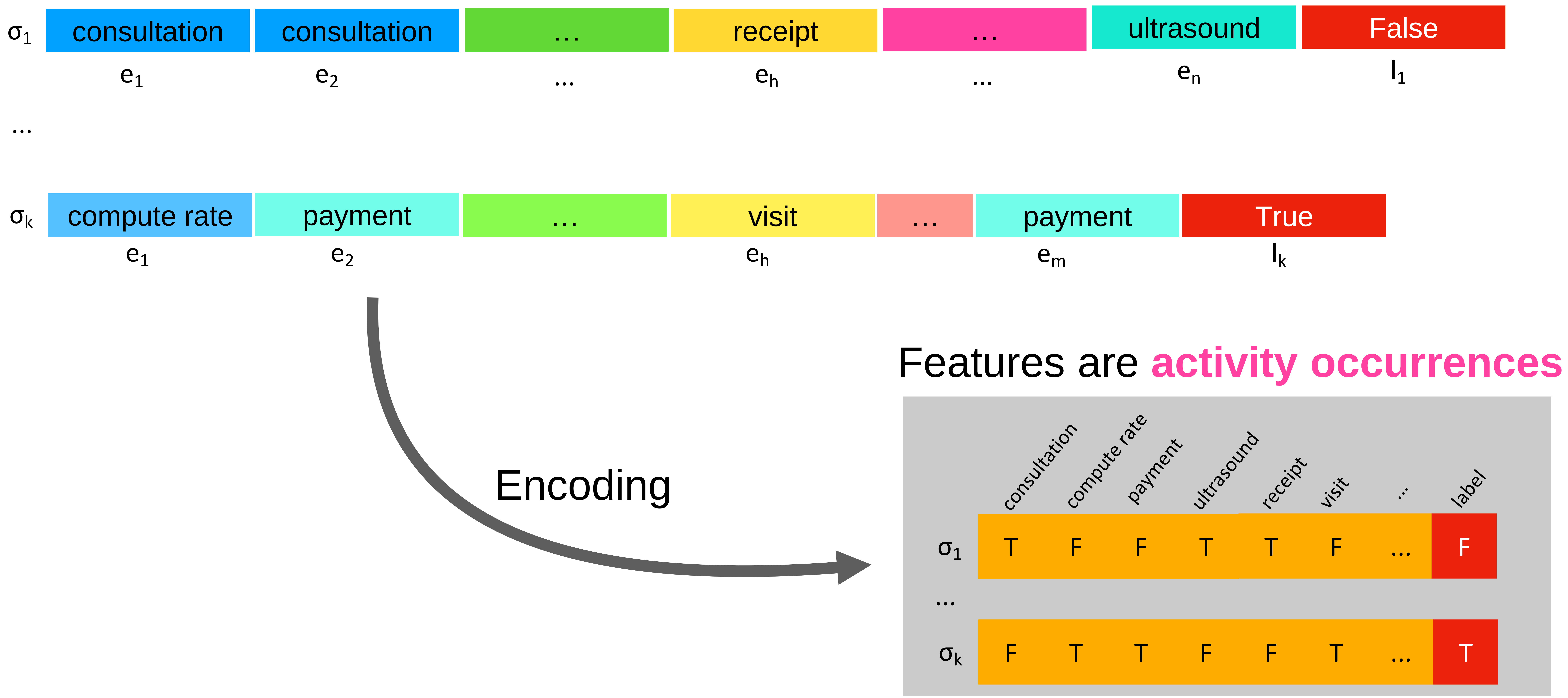
Predictive model
all prefix lengths

The “traditional” pipeline

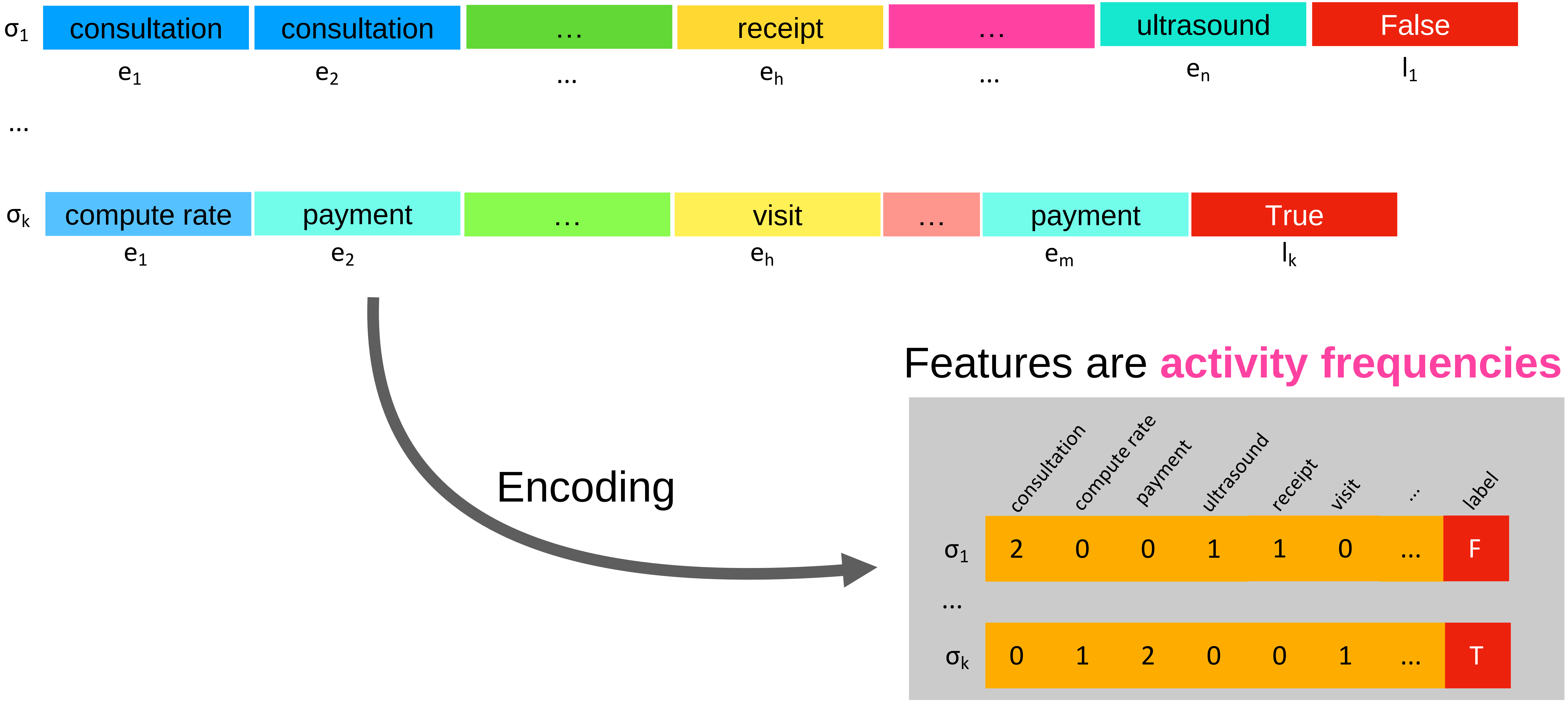
The “traditional” pipeline

A. Leontjeva, R. Conforti, C. Di Francescomarino, M. Dumas, F. M. Maggi: **Complex Symbolic Sequence Encodings for Predictive Monitoring of Business Processes**. BPM 2015: 297-313

Boolean encoding

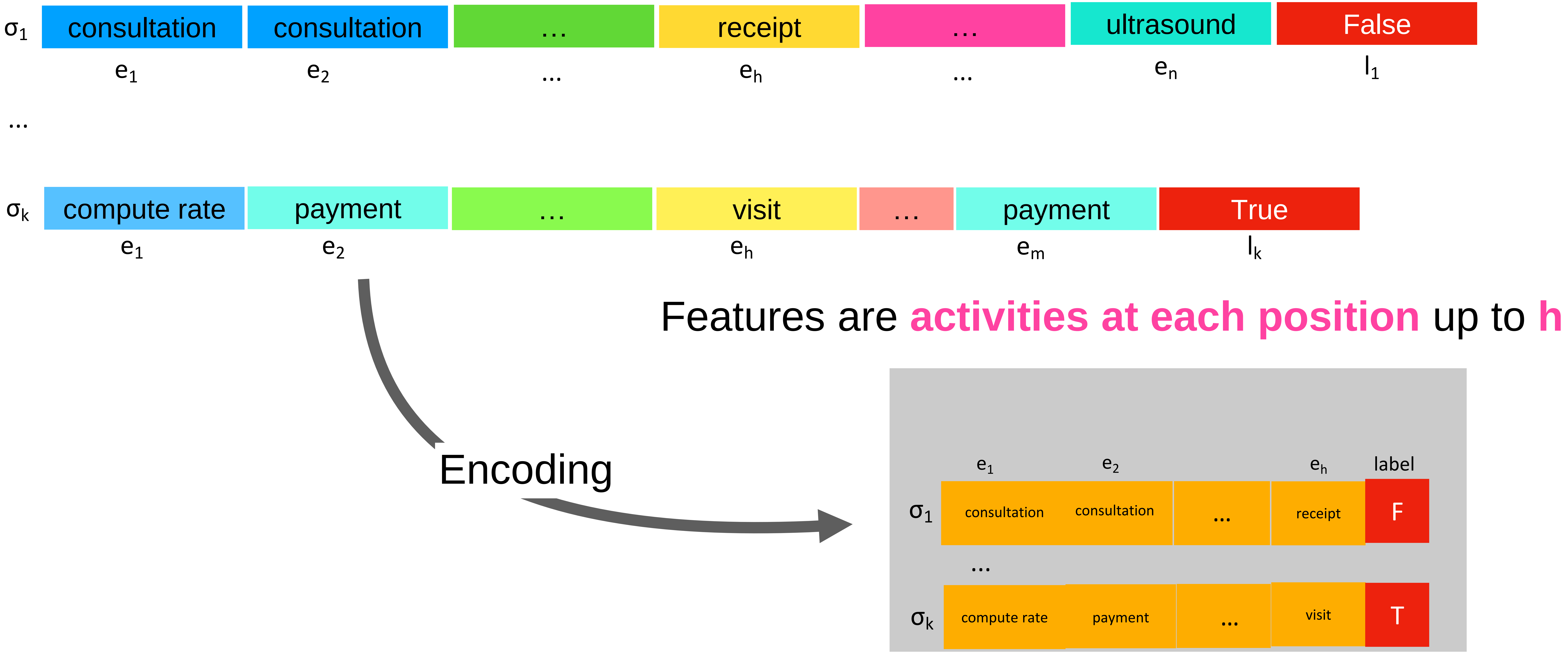


Frequency encoding



Simple index encoding

How to consider the ordering



Index latest-payload encoding

How to start considering data

	e_1	e_2		e_h		e_m	l_k
σ_1	consultation *33* *clinic*	consultation *33* *lab*	… …	receipt *33* *100EUR*	… …	ultrasound *33* *radiology*	False
…							
σ_k	compute rate *56* *clinic*	payment *56* *admin*	… …	visit *56* *lab3*	… …	payment *56* *admin*	True

As simple index + **trace attribute values** + **datapayload values at h**

Encoding

	age	e_1	e_2		e_h	depth_h	label
σ_1	*33*	consultation	consultation	...	receipt	*100EUR*	F
		...					
σ_k	*56*	compute rate	payment	...	visit	*lab3*	T

Complex index encoding

The full monty

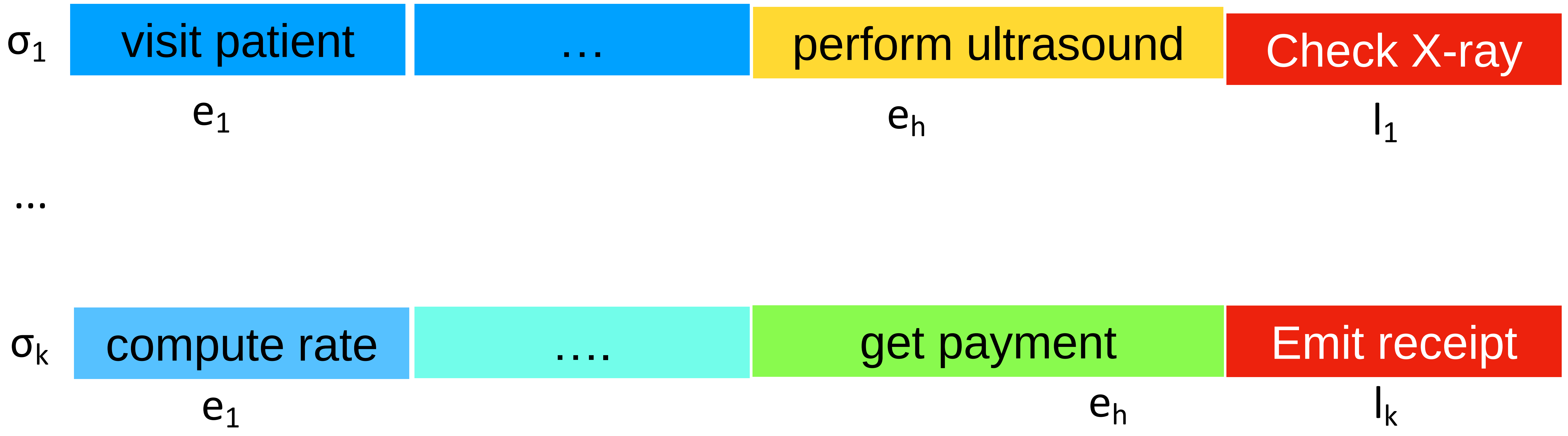
	e_1		e_h		e_m	l_k
σ_1	consultation *33* *clinic*	… …	receipt *33* *100EUR*	… …	ultrasound *33* *radiology*	False
…						
σ_k	compute rate *56* *clinic*	… …	visit *56* *lab3*	… …	payment *56* *admin*	True

Features are **activities and data payload values** at each position up to h,
+ **trace attribute values**

Encoding

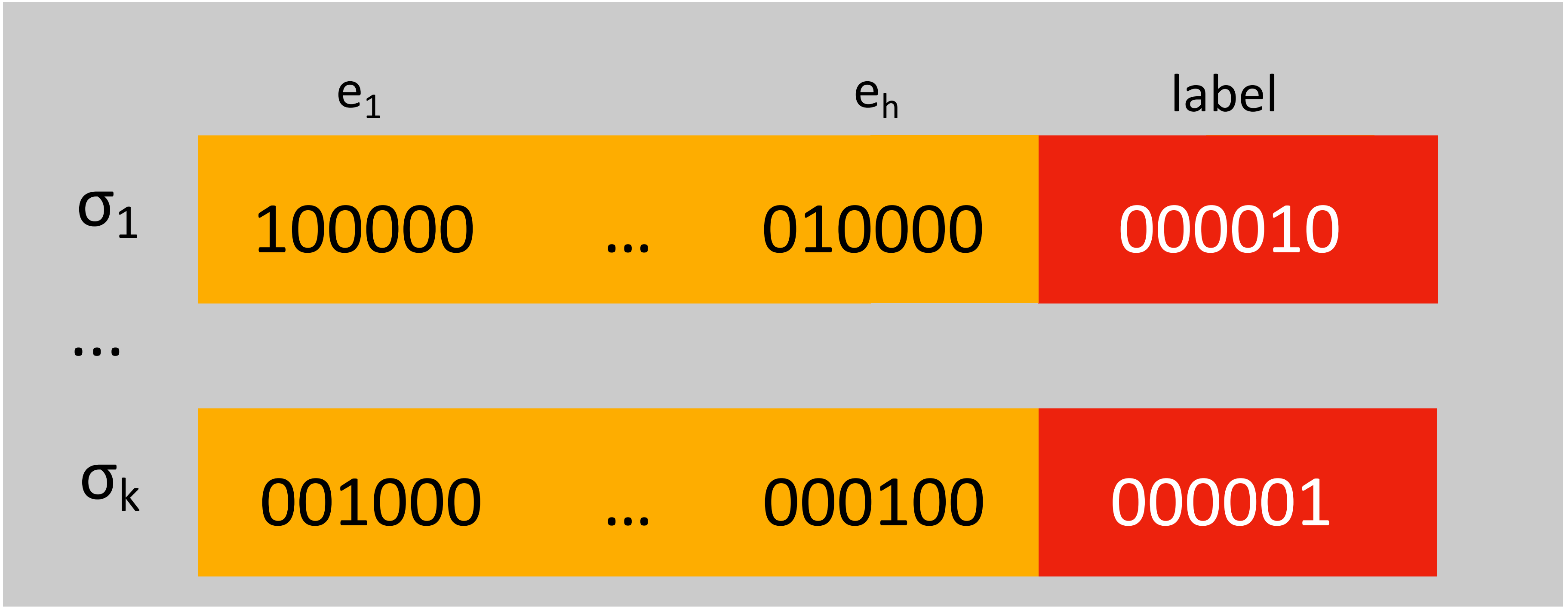
	age	e_1		e_h	$depth_1$		$depth_h$	label
σ_1	*33*	consultation	…	receipt	*clinic*	…	*100EUR*	F
		…						
σ_k	*56*	compute rate	…	visit	*clinic*	…	*lab3*	T

One-hot encoding (for next activity)



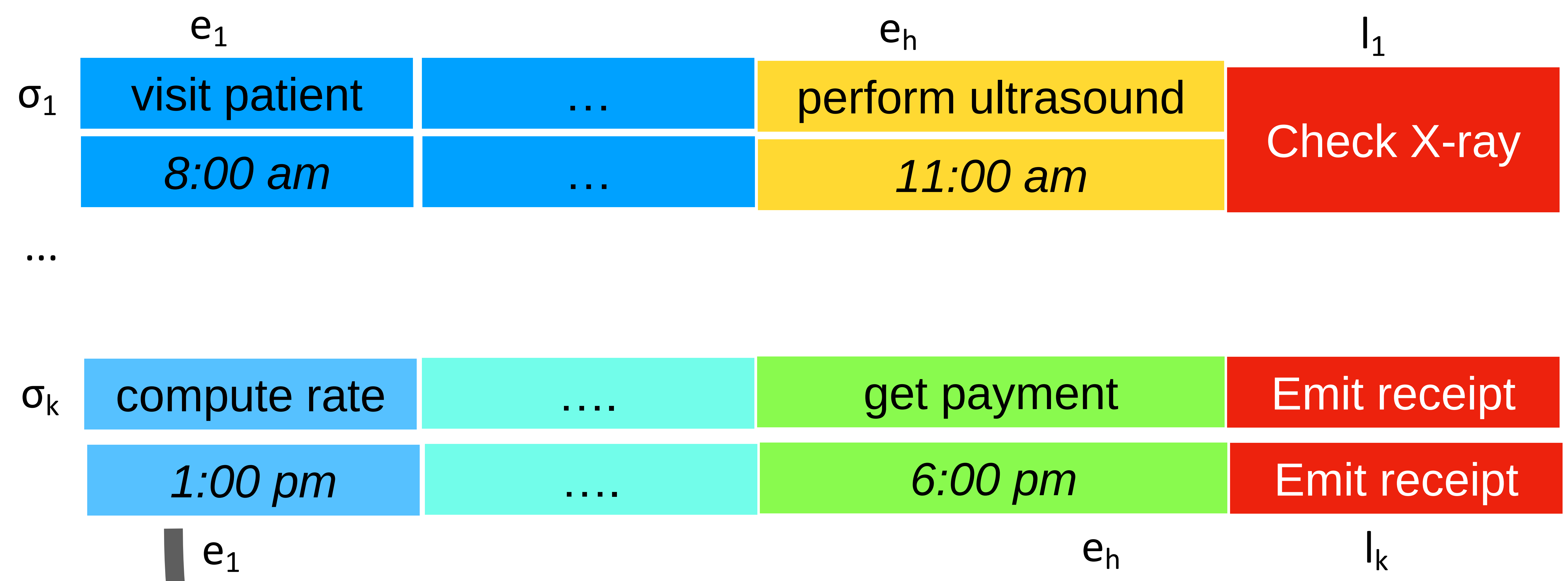
Encoding

Features and label are **binary numbers** from an ordered alphabet up to h.



Alphabet	Index
Visit patient	1
Perform ultrasound	2
Compute rate	3
Get payment	4
Check X-ray	5
Emit receipt	6

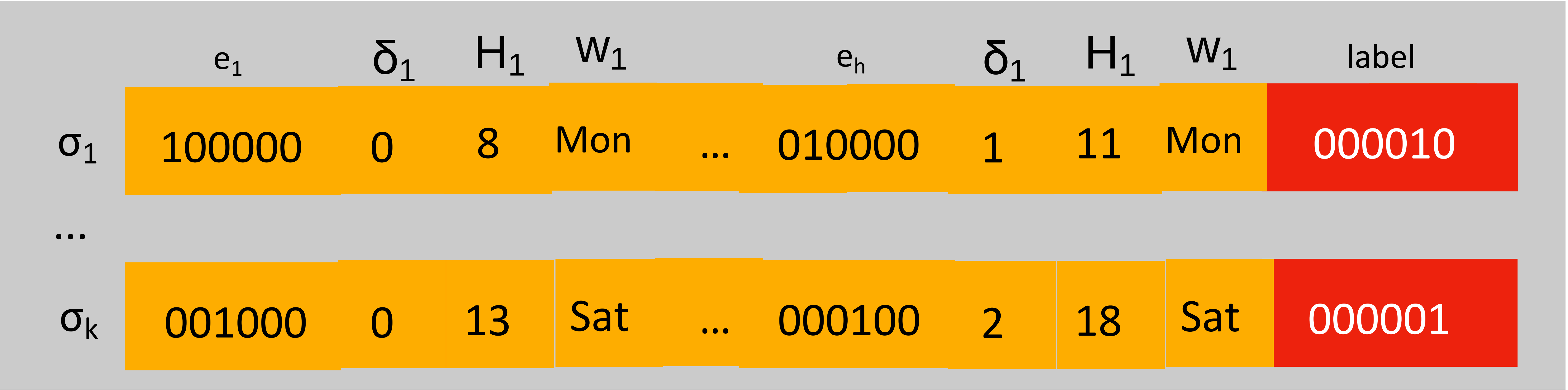
Plus temporal features!



Alphabet	Index
Visit patient	1
Perform ultrasound	2
Compute rate	3
Get payment	4
Check X-ray	5
Emit receipt	6

As one hot encoding plus **temporal features**

Encoding

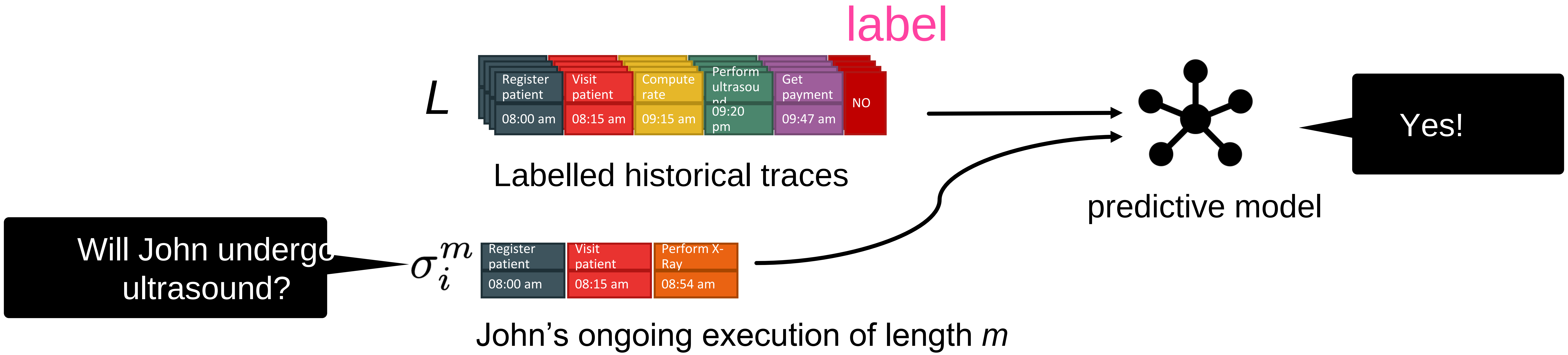


The "traditional" pipeline

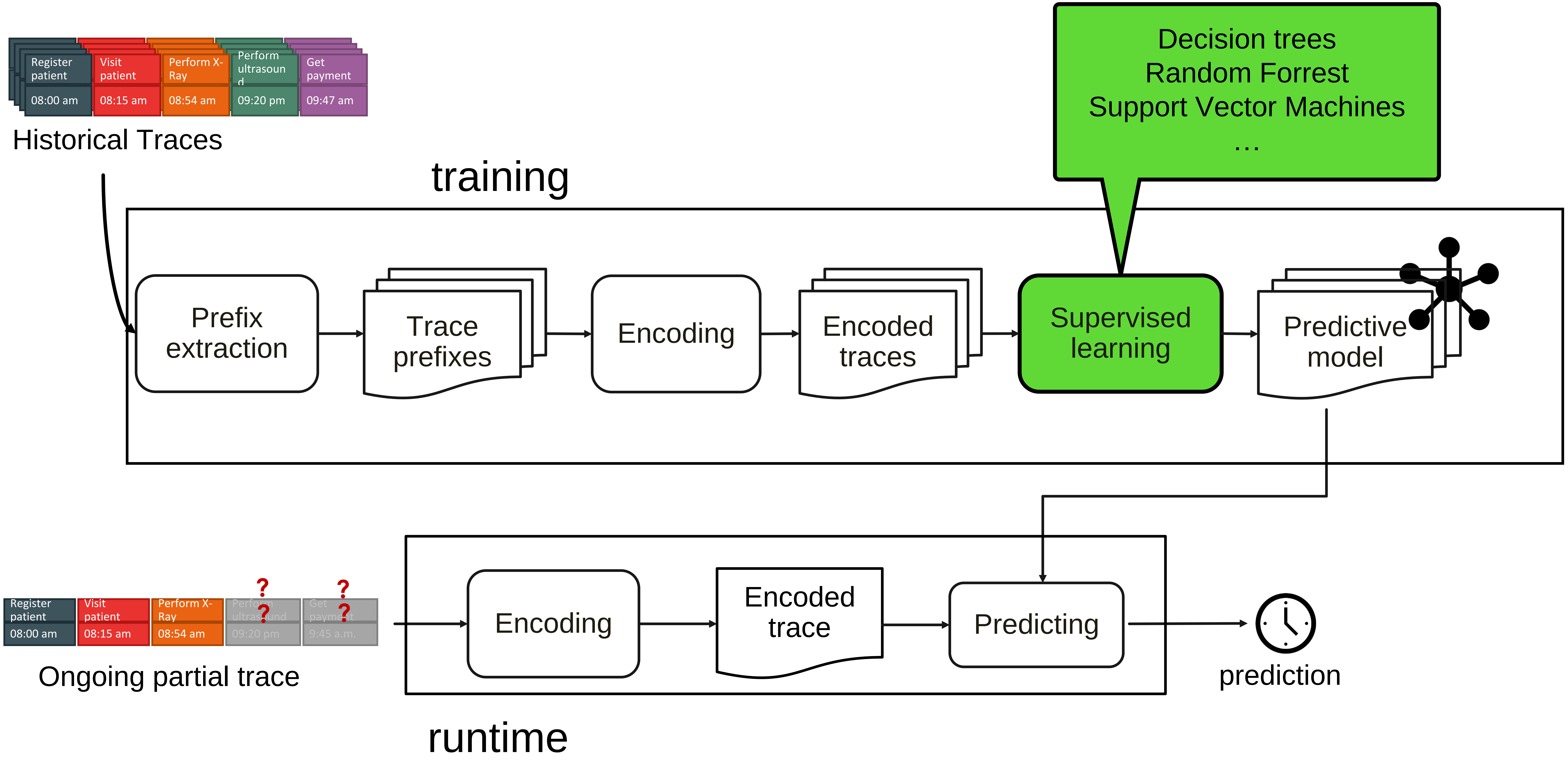
Outcome-based predictions

The idea

- Prediction of **categorical values** (e.g., true/false, good/average/bad)
- The **label** is a categorical value
- Given an event log *L* and an ongoing execution σ_i^m of length *m*, we want to **learn a function** $f_c(L, \sigma_i^m) = \overline{label_i}$ as close as possible to the actual label



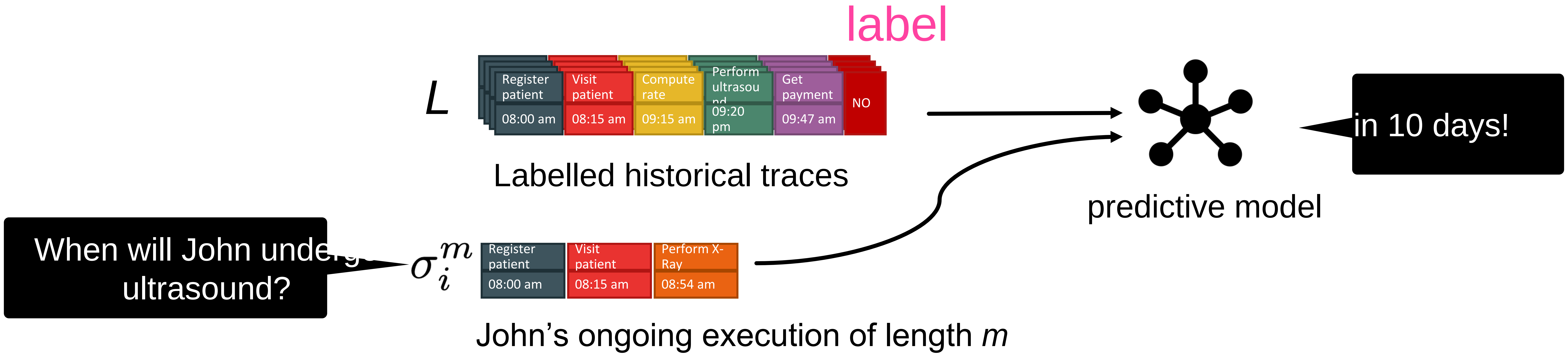
Classification-based approaches



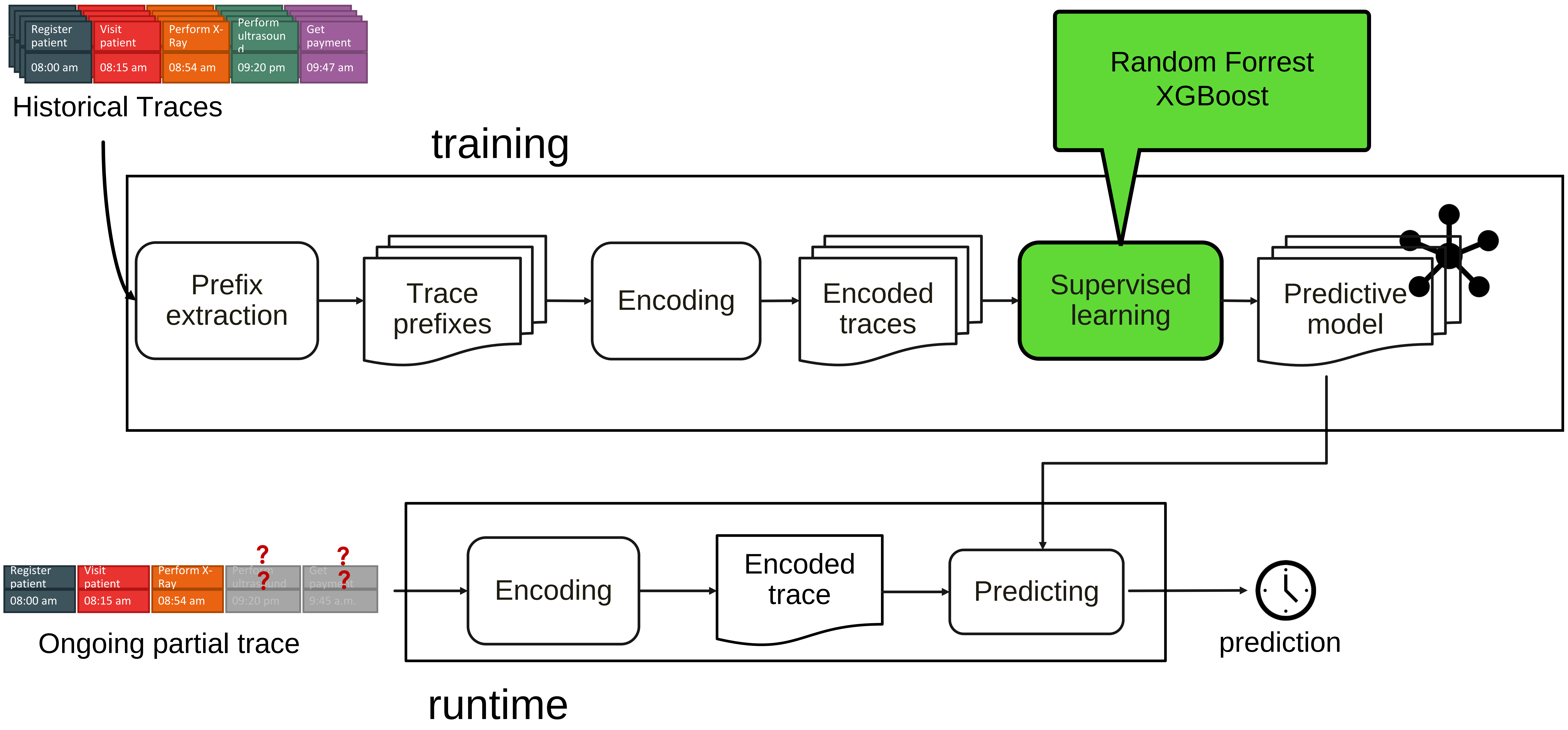
Numerical –value predictions

The idea

- Prediction of **numerical values** (e.g., the remaining time, the cost)
- The **label** is a numerical value
- Given an event log *L* and an ongoing execution σ_i^m of length *m*, we want to **learn a function** $f_n(L, \sigma_i^m) = \overline{label_i}$ as close as possible to the actual label



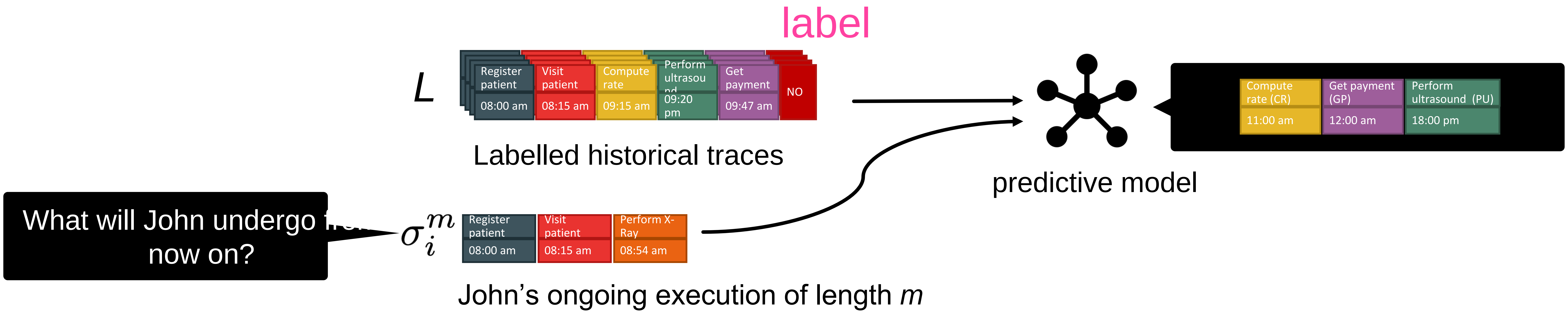
Regression-based approaches



Next-event predictions

The idea

- Prediction of **future events** (event class or data payload)
- Usually approaches first learn a function f_{1a} that given the first *m* events predicts the next event class and then iteratively predict the suffix until the last event ω.



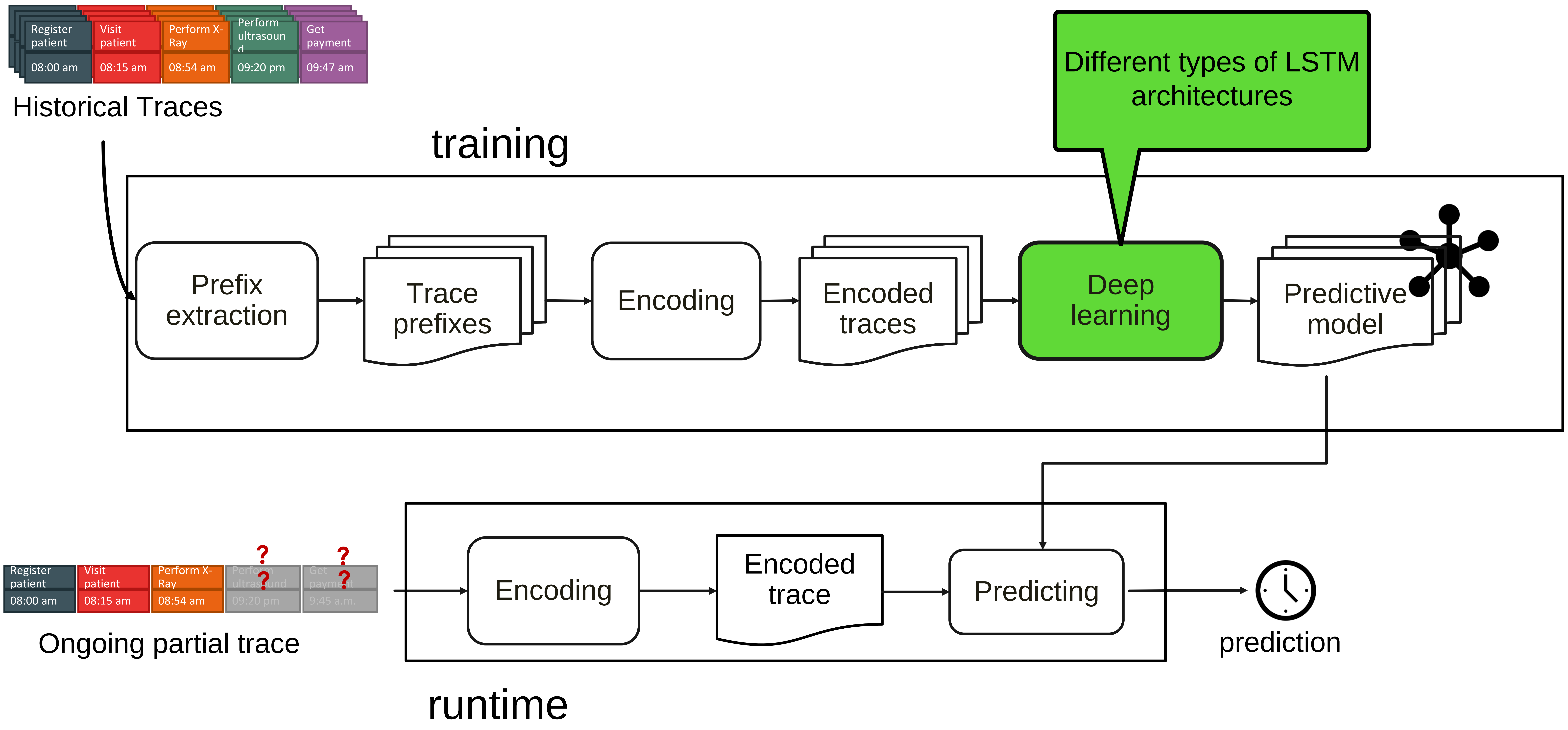
Next-event predictions

The idea

- Given an event log L and an ongoing execution σ_i^m of length m, we want to **learn a function** as the one below, as close as possible to the actual sequence of activities.

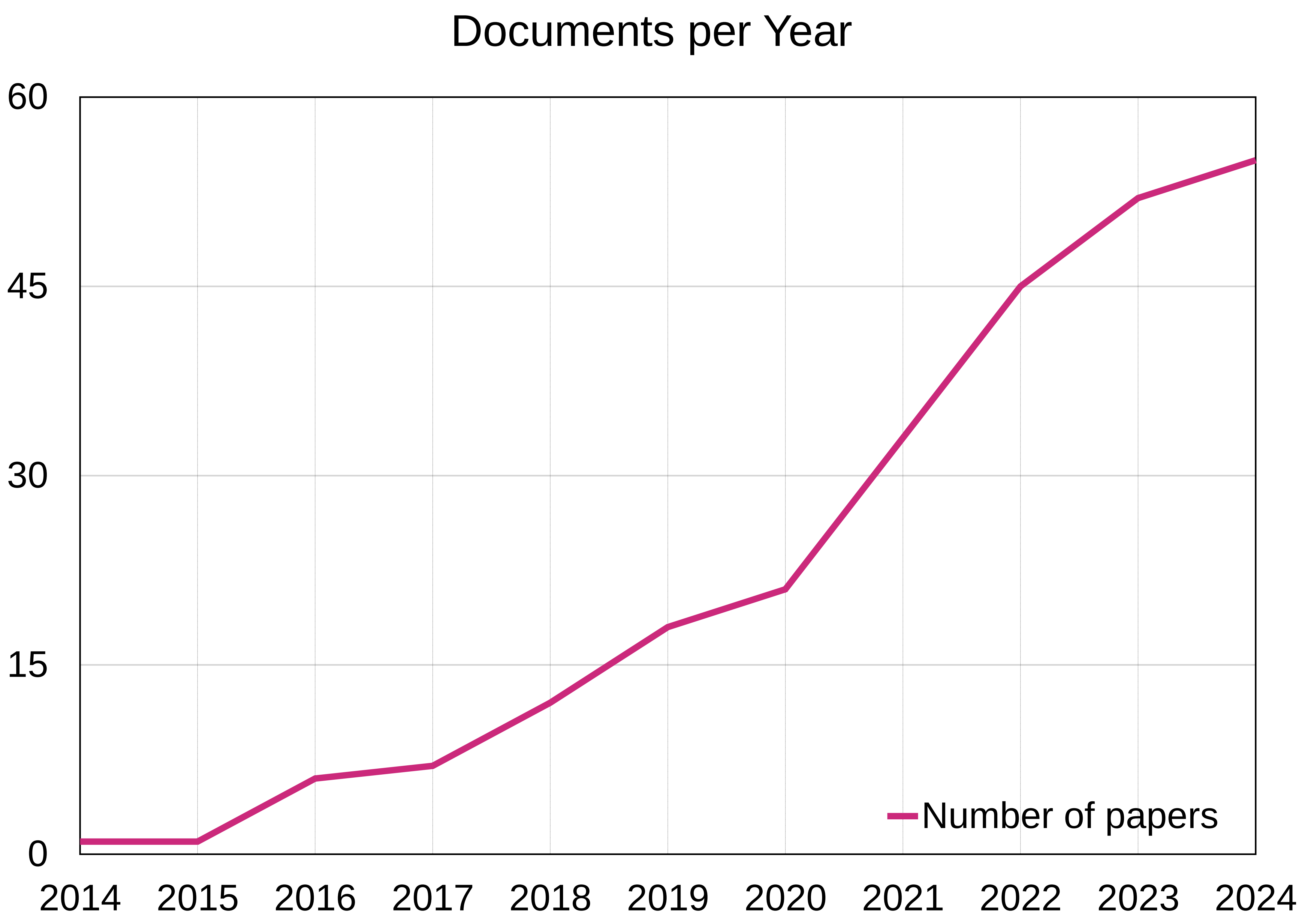
$$f_{sa}(L, \sigma_i^m) = \begin{cases} f_{1a}(\sigma^m) & \text{if } f_{1a}(L, \sigma_i^m) = \omega \\ f_{sa}(L, < e_1, e_2, ..., e_m, e >), & \\ \qquad \text{with } e\text{'s event class computed as } f_{1a}(L, \sigma_i^m) & \text{otherwise} \end{cases}$$

LSTM-based approaches



Summing up

- A healthy field
- Scopus: TITLE-ABS-KEY (“predictive process monitoring”)



Fabrizio Maria Maggi, Chiara Di Francescomarino, Marlon Dumas, Chiara Ghidini: Predictive Monitoring of Business Processes. CAiSE 2014: 457-472

What next in ML-driven operational support?

Chiara's personal view on what should be there

This Photo by Unknown Author is licensed

Half Time (technical) Show

References

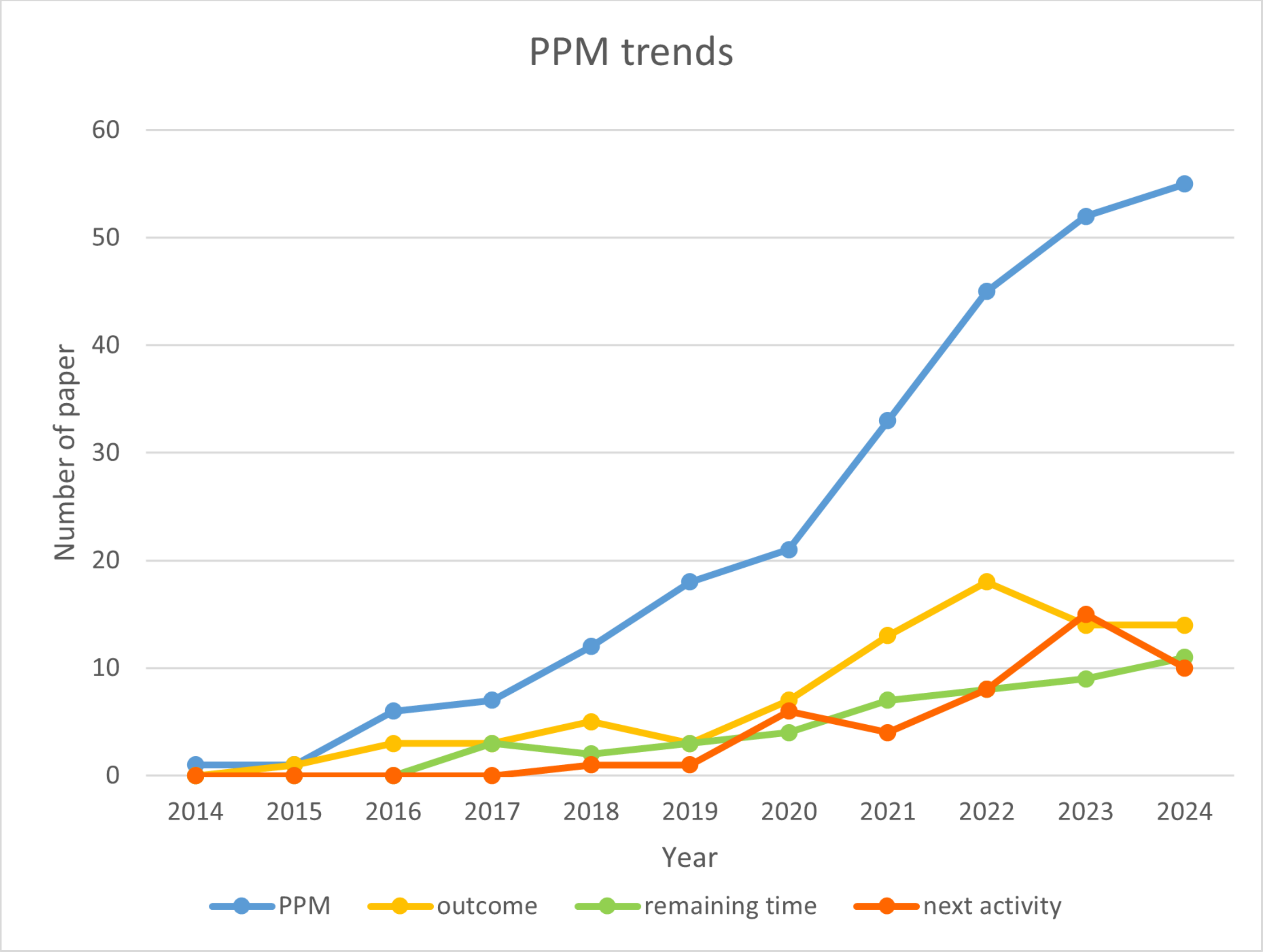
F. Taymouri, M. La Rosa, S. M. Erfani, "A Deep Adversarial Model for Suffix and Remaining Time Prediction of Event Sequences." SDM (2021)

I. Ketykó, F. Mannhardt, M. Hassani, B. F. van Dongen, "What averages do not tell: predicting real life processes with sequential deep learning." SAC22 (2022)

B. R. Gunnarsson, S. v. Broucke, J. De Weerdt, "A Direct Data Aware LSTM Neural Network Architecture for Complete Remaining Trace and Runtime Prediction." IEEE Transactions on Services Computing, vol. 16, no. 4, (2023)

M. Camargo, M. Dumas, O. González-Rojas. "Learning Accurate LSTM Models of Business Processes." BPM19 (2019).

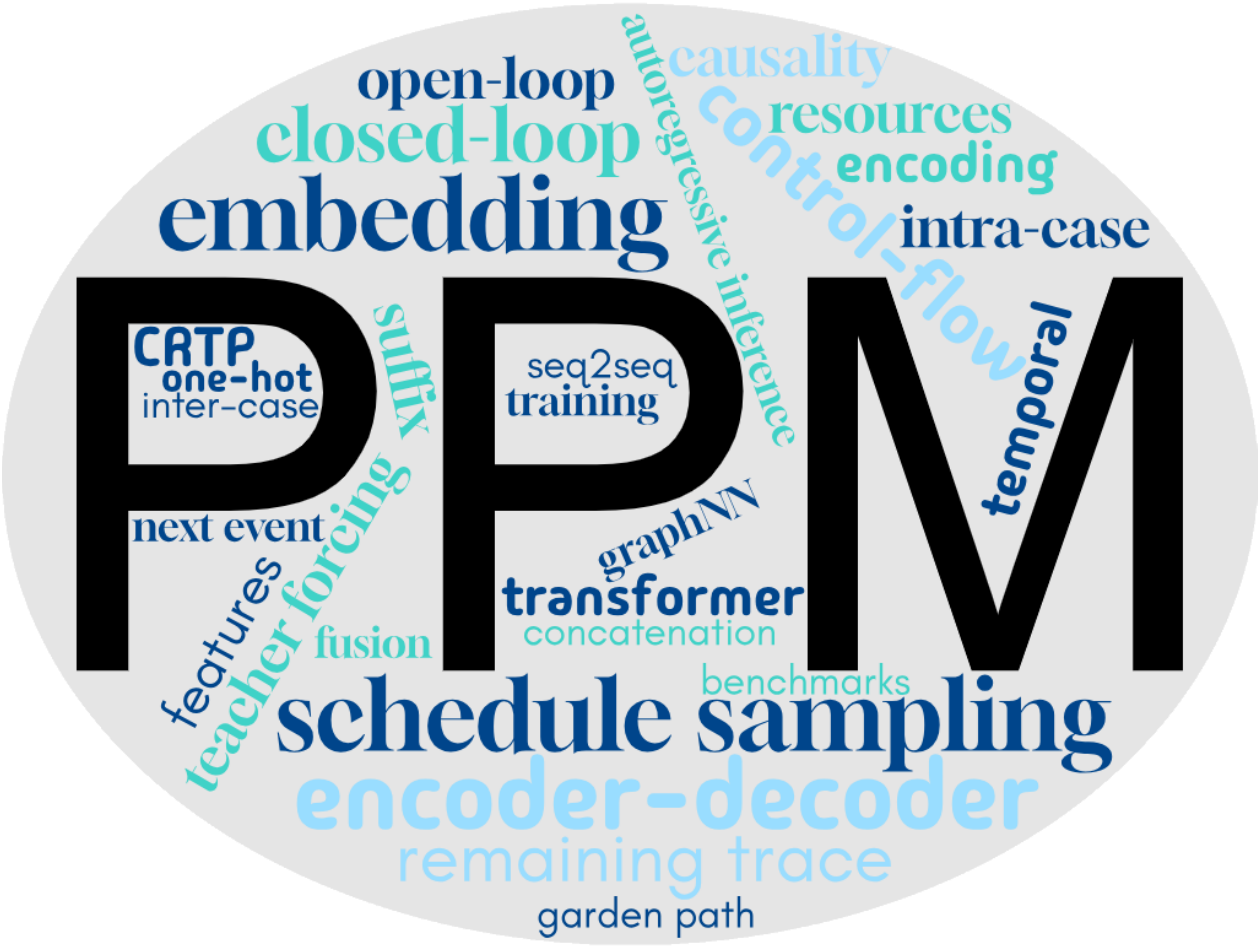
What are the trends in PPM?



- *LUPIN: A LLM Approach for Activity Suffix Prediction in Business Process Event Logs.* V. Pasquadibisceglie, A. Appice and D. Malerba.

- *SuTraN: an Encoder-Decoder Transformer for Full Context-Aware Suffix Prediction of Business Processes.* B. Wuyts, S. Vanden Broucke and J. De Weerdt.

Next event and suffix prediction



Where to seek new techniques ML techniques?

NLP

- Sequential data
- Categorical features
- Temporal dimension
- Causal constraints
- Multimodality
- Inter-sample dependency

PPM

Where to seek new techniques ML techniques?

PPM

Time-series

- Sequential data
- Categorical features
- Temporal dimension
- Causal constraints
- Multimodality
- Inter-sample dependency

Where to seek new techniques ML techniques?

NLP

Graph-NN

PPM

Vision

Time-series

Multimodal models

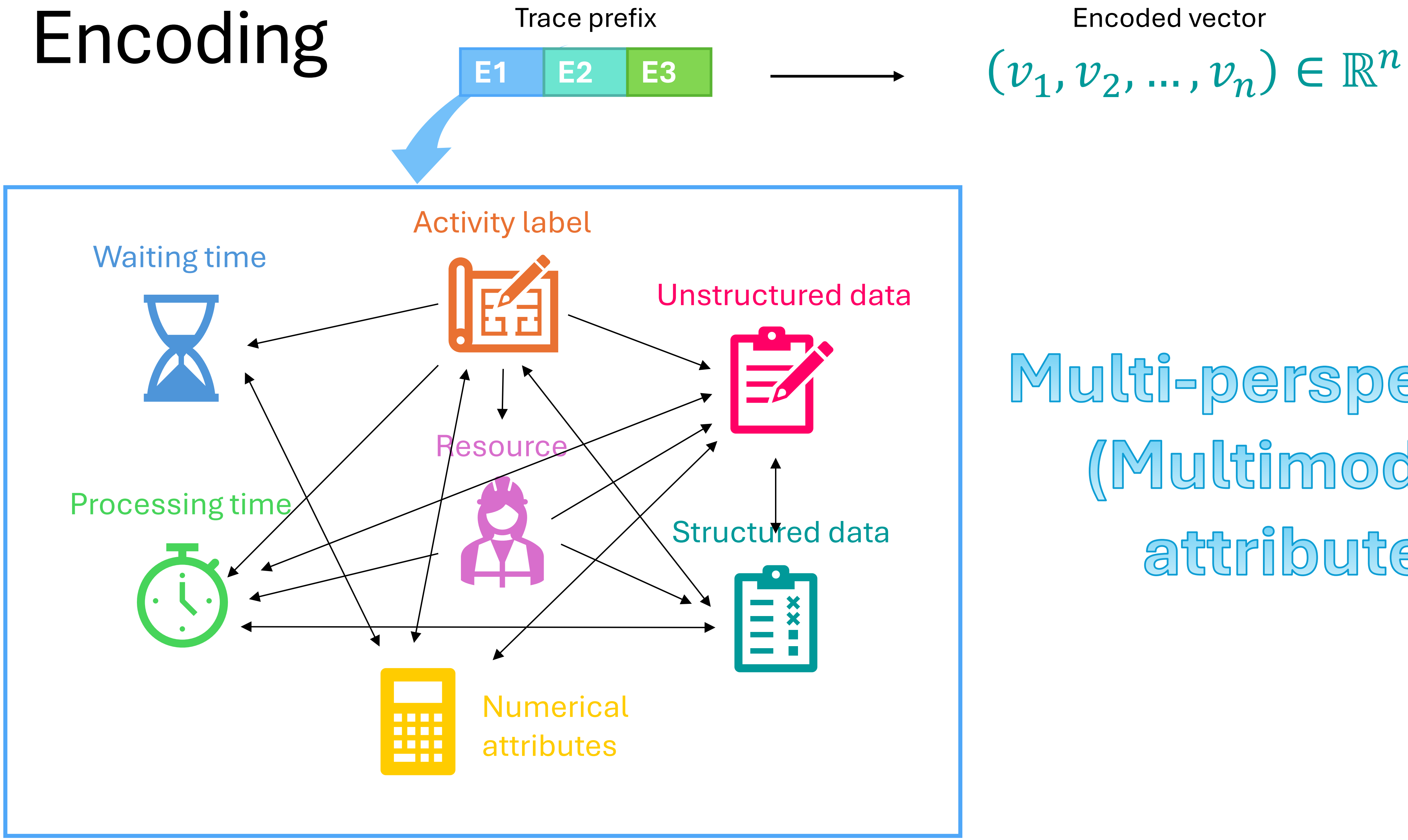
Where can we apply ML techniques in PPM?

- Preprocessing
- Encoding of traces
- Training and inference
- Evaluation

~~**Encoding**~~

Representation Learning

Encoding



Multi-perspective (Multimodal) attributes

Categorical features

Given the set of categorical (or nominal) feature values
$A = \{a_1, \dots, a_N\}$

- **One-hot vectors:** $a_j \to \vec{v}_j = \delta_{i,j} = (0, \dots, \overset{j\text{-th}}{\overset{\downarrow}{0,1,0}}, \dots, 0) \in \mathbb{R}^N$ Introduces many dimensions
- **Label encoding:** $a_j \to j/(N) \in \mathbb{R}$ Introduce a fictitious ordering
- **Embedded vectors:** $a_j \to V(a_j) \in \mathbb{R}^d, 1 < d < N$

Popular choice: many possibilities, hard embeddings, soft, pretrained, end-to-end

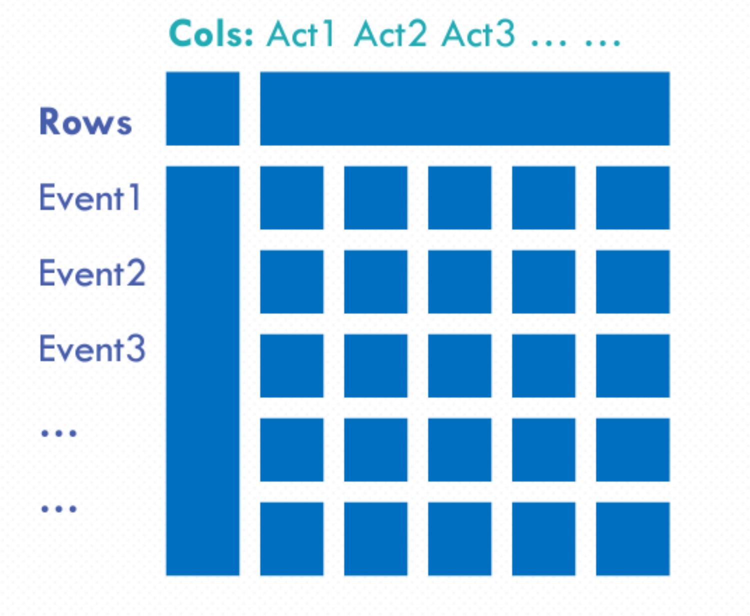
Multimodal attributes fusion

How do we put together the different perspectives?

- **Early fusion:** features vector concatenation
- **Cross-feature embedding:** embed the combination of multiple features: e.g. Activity + Role
- **Vision:** Represent traces as 2D images to leverage CNNs
- **Multimodal models:** tensor fusion, multimodal attention

From **Vision**

Represent traces as 2D images to leverage CNNs

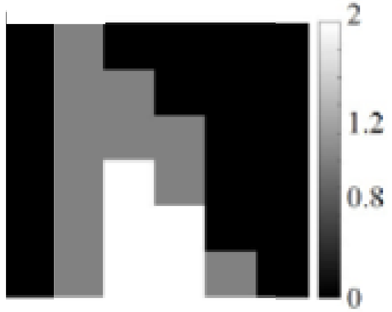


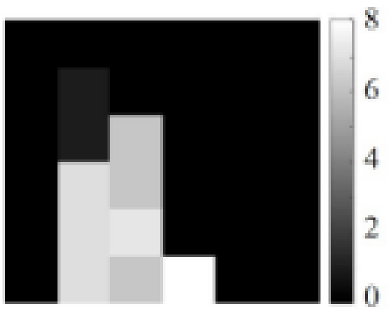
Activity Channel-Matrix:

R	ET	CT	D	RR	AR
0	1	0	0	0	0
0	1	1	0	0	0
0	1	1	1	0	0
0	1	2	1	0	0
0	1	2	2	0	0
0	1	2	2	1	0

Performance Channel-Matrix

R	ET	CT	D	RR	AR
0	0	0	0	0	0
0	0,9611	0	0	0	0
0	0,9611	6,1736	0	0	0
0	7,0002	6,1736	0	0	0
0	7,0020	7,1736	0	0	0
0	7,0002	6,1732	8,0111	0	0

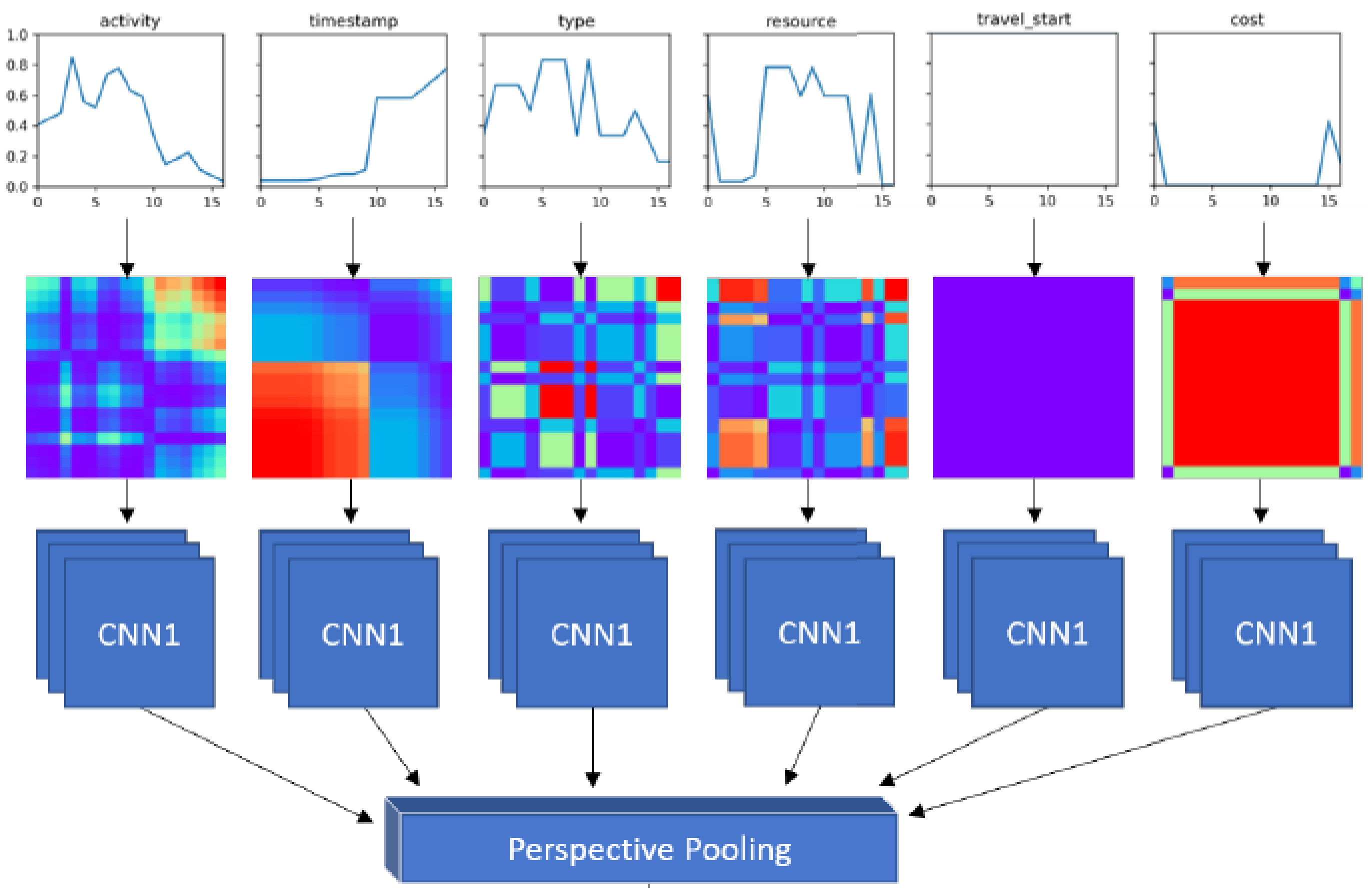




V. Pasquadibisceglie, A. Appice, G. Castellano, D. Malerba, "Using Convolutional Neural Networks for Predictive Process Analytics", ICPM19 (2019)

From Multivariate Time Series

Gramian Angular Fields (GAF)

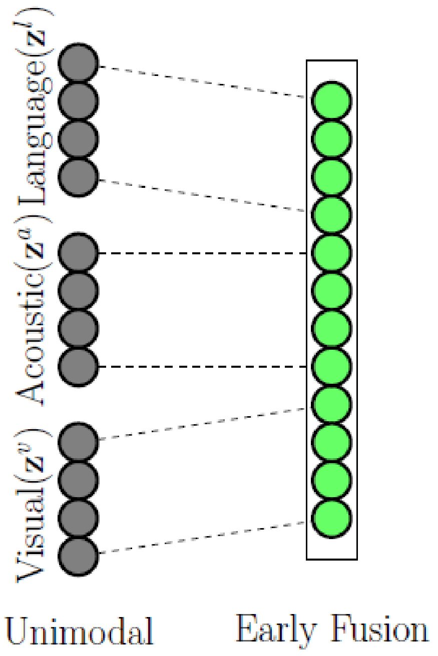


P. Pfeiffer, J. Lahann, P. Fettke, “Multivariate Business Process Representation Learning Utilizing Gramian Angular Fields and Convolutional Neural Networks.” BPM21 (2021)

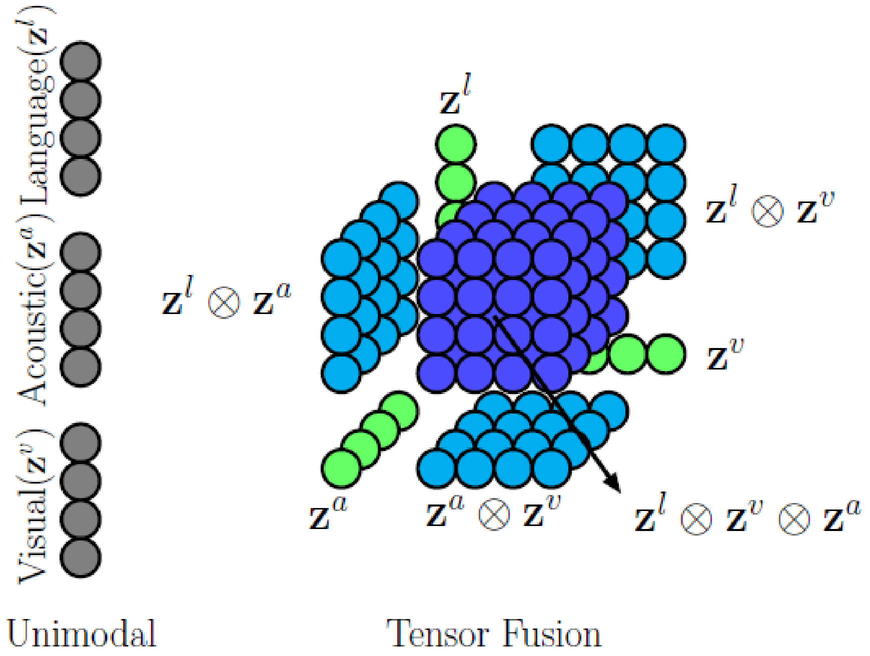
Other inspirations from Multimodal models

3 features: Language, Visual, Audio (L,V,A)

Early Fusion



Tensor Fusion

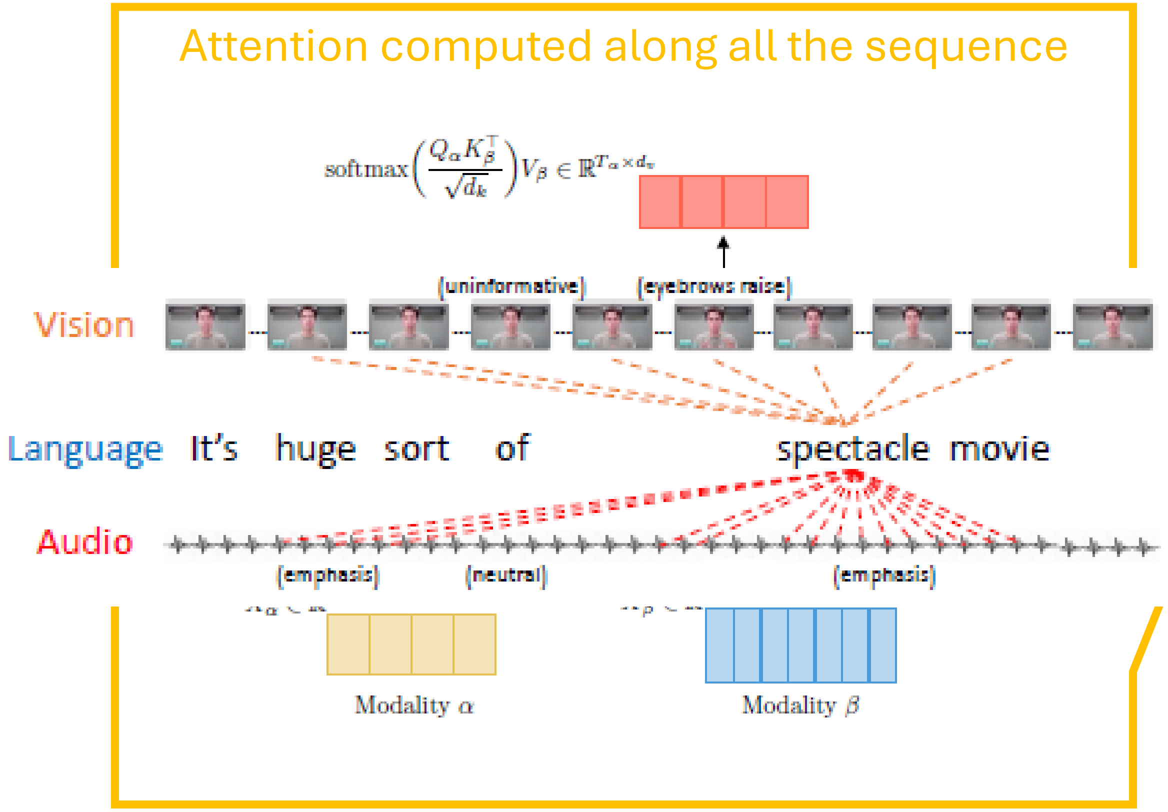


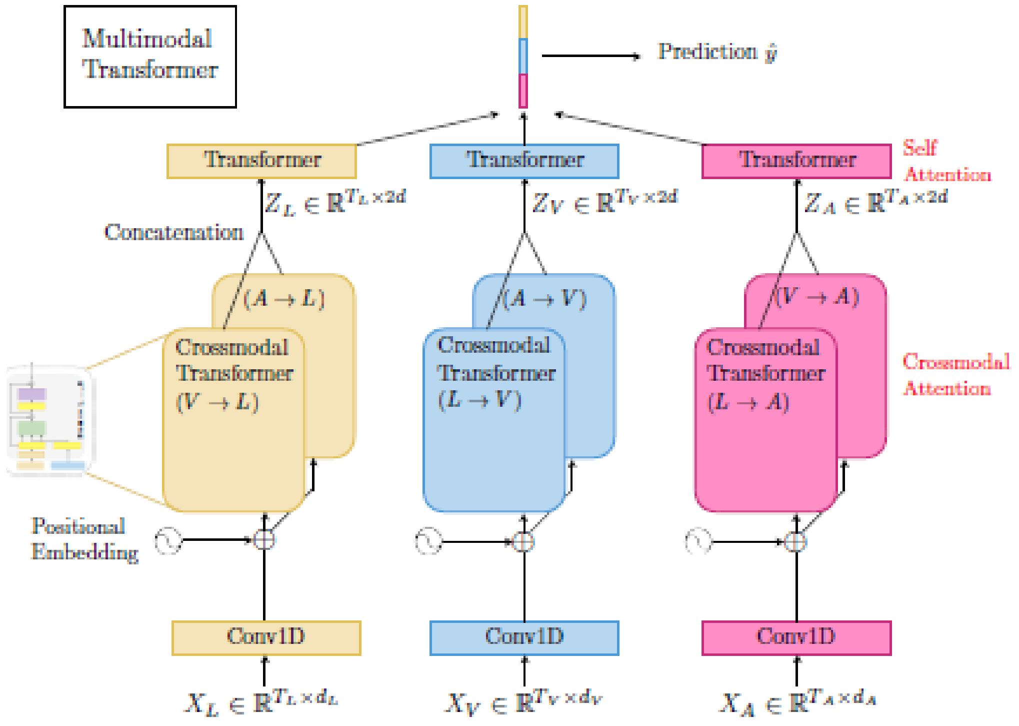
Outer product of different unimodal features

A. Zadeh, M. Chen, S. Poria, E. Cambria, L.-P. Morency, “Tensor Fusion Network for Multimodal Sentiment Analysis.” EMNLP (2017)

Other inspirations from **Multimodal models**

Crossmodal attention: multimodal transformer





Y.-H. H. Tsai, S. Bai, P. P. Liang, J. Z.Kolter, L-P Morency, R. Salakhutdinov, « Multimodal Transformer for Unaligned Multimodal Language Sequences." ACL (2019)

Age of the Exploration

Training and inference

Next event and suffix prediction

Neural architectures

LSTM

- J. Evermann, J.-R. Rehse, and P. Fettke, “A deep learning approach for predicting process behaviour at runtime,’’ BPM17, (2017).
- N. Tax, I. Verenich, M. La Rosa, M. Dumas, “Predictive business process monitoring with LSTM neural networks,’’CAiSE17, (2017).
- M. Camargo, M. Dumas, O. G. Rojas, “Learning accurate LSTM models of business processes,’’ BPM19, (2019).
- ....

TRANSFORMER

- Z. A. Bukhsh, A. Saeed, and R. M. Dijkman, “ProcessTransformer: Predictive Business Process Monitoring with Transformer Network,” 2021
- G. Rivera Lazo,R. Ñanculef, “Multi-attribute Transformers for Sequence Prediction in Business Process Management,” in Discovery Science, 2022
- I. Ketykó, F. Mannhardt, M. Hassani, B. F. van Dongen, “What averages do not tell: predicting real life processes with sequential deep learning.” SAC22 (2022)
- ....

From next event to suffix

N. Tax, I. Verenich, M. La Rosa, M. Dumas, «Predictive Business Process Monitoring with LSTM Neural Networks.» CAiSE 2017.

Prefix

e1	e2	e3	→ NN →	e4

Next event

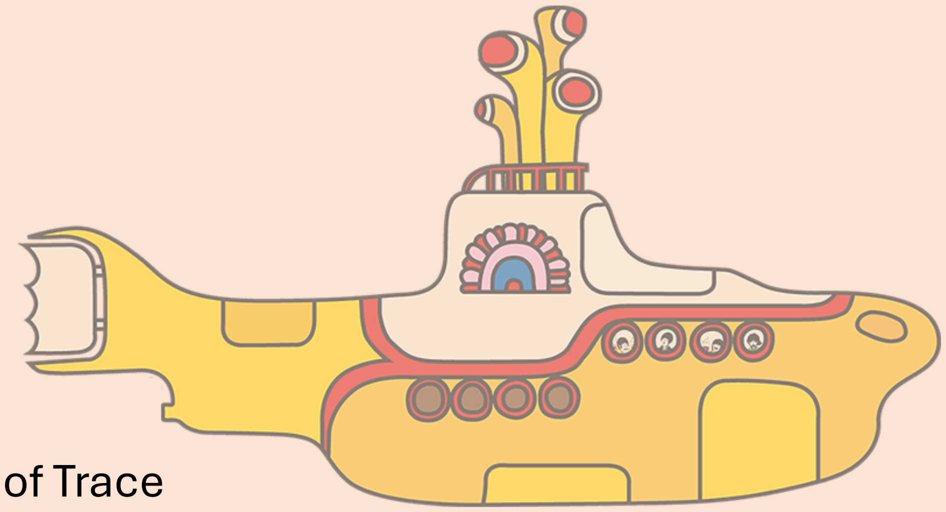
Hallucination mechanism

e1	e2	e3	e4	→ NN →	e5

⋮

e1	e2	e3	e4	e5	e6	→ NN →	EOS

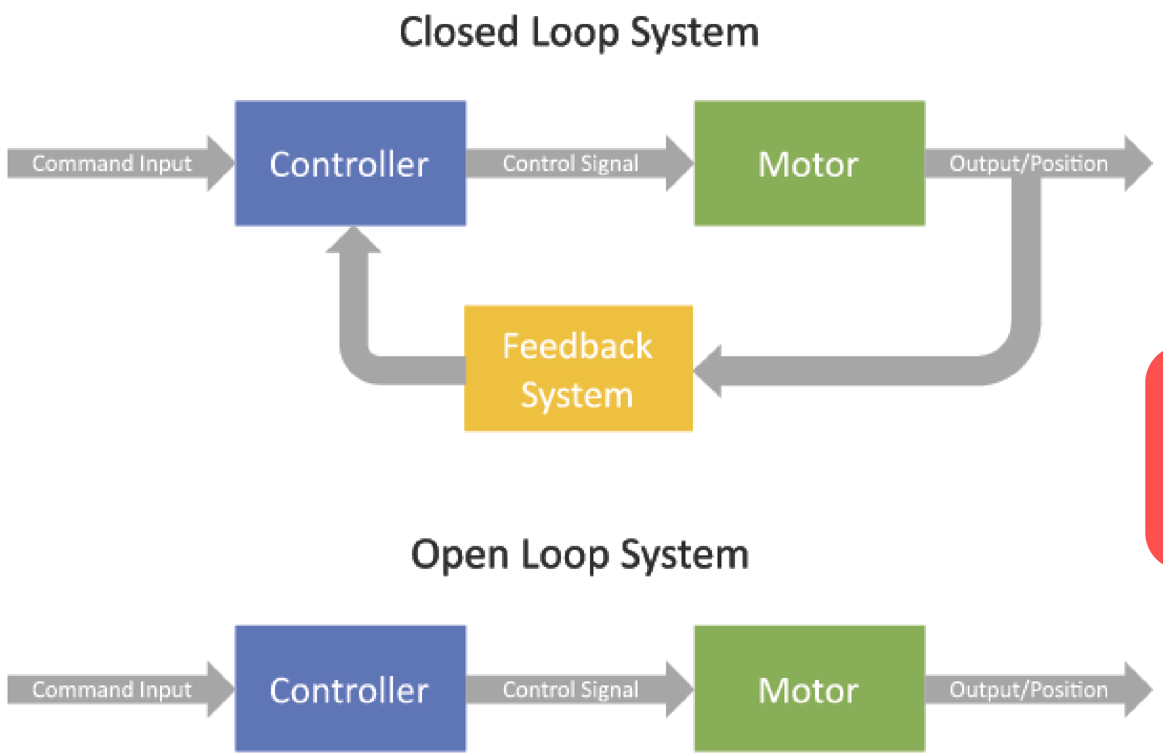
End of Trace



Open-loop training and closed-loop inference

Challenging in scenarios involving temporal dependencies or sequential decision-making

From control systems



Suffix prediction:
Single event prediction +
Autoregressive inference

Training: (on single event prediction) the next event is conditioned on the ground truth of previous events

Feedback Dependency, Error Accumulation

Inference: the next event is conditioned on previously predicted events

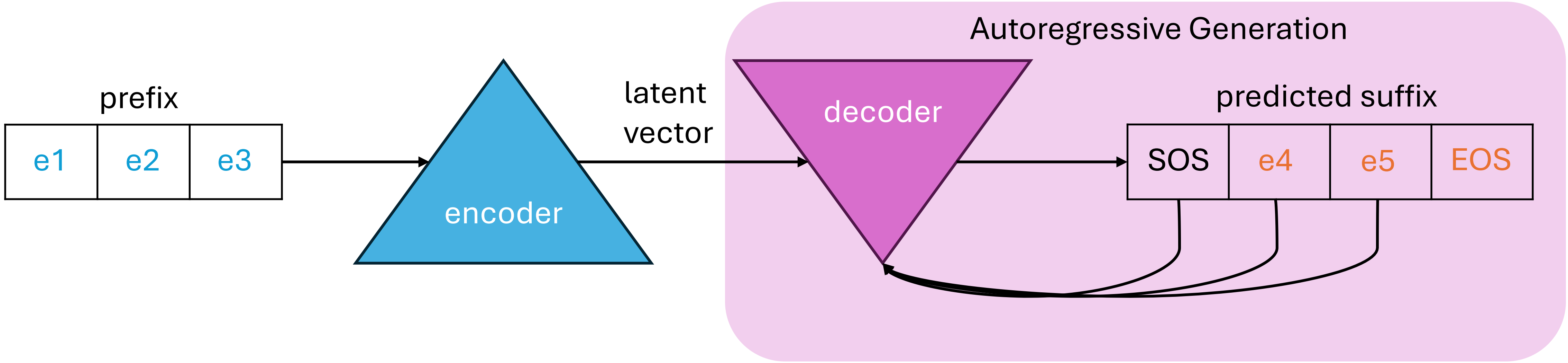
Open-loop training and closed-loop inference

Challenging in scenarios involving temporal dependencies or sequential decision-making

Encoder-decoder architecture (from **NLP** Seq2Seq)

- **Encoder:** encodes the prefix in a latent space and pass it to the decoder
- **Decoder:** autoregressively (AR) generates the trace

Loss is computed ultimately between ground truth suffix and the predicted one



I. Sutskever, O. Vinyals, Q. V. Le, “Sequence to sequence learning with neural networks”, NIPS14, (2014)

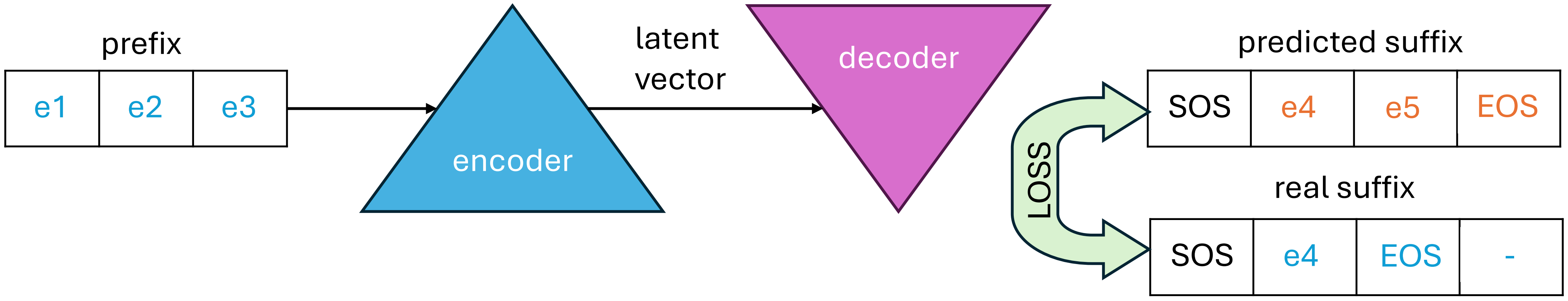
Open-loop training and closed-loop inference

Challenging in scenarios involving temporal dependencies or sequential decision-making

Encoder-decoder architecture (from **NLP** Seq2Seq)

- **Encoder:** encodes the prefix in a latent space and pass it to the decoder
- **Decoder:** autoregressively (AR) generates the trace

Loss is computed ultimately between ground truth suffix and the predicted one

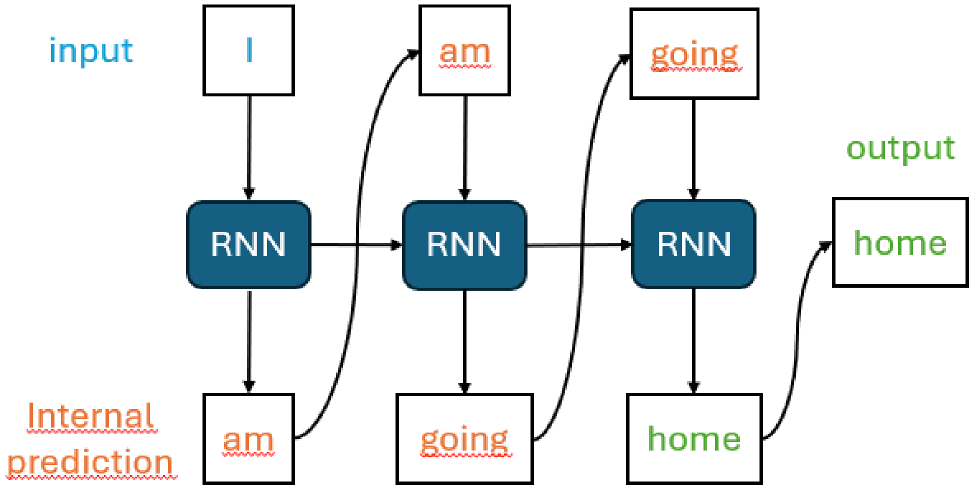


I. Sutskever, O. Vinyals, Q. V. Le, "Sequence to sequence learning with neural networks", NIPS14, (2014)

Error accumulation

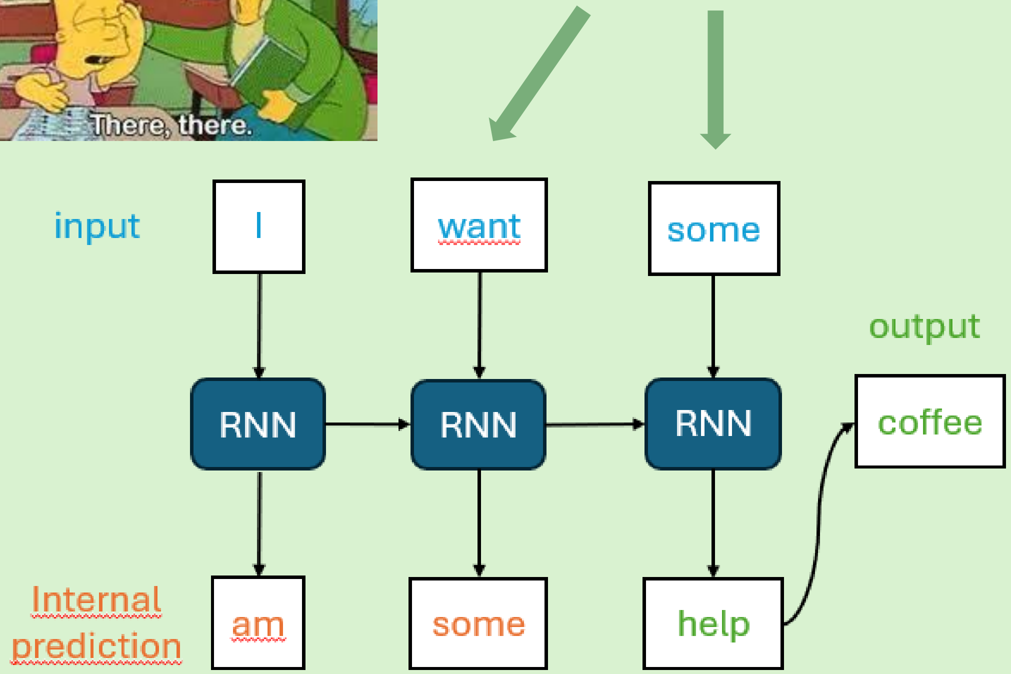
The error is accumulated during training

Example: “I want some ice-cream”



Teacher forcing

Feed ground truth values



Scheduled Sampling: Introduce gradually to the model its own prediction

Garden path problem

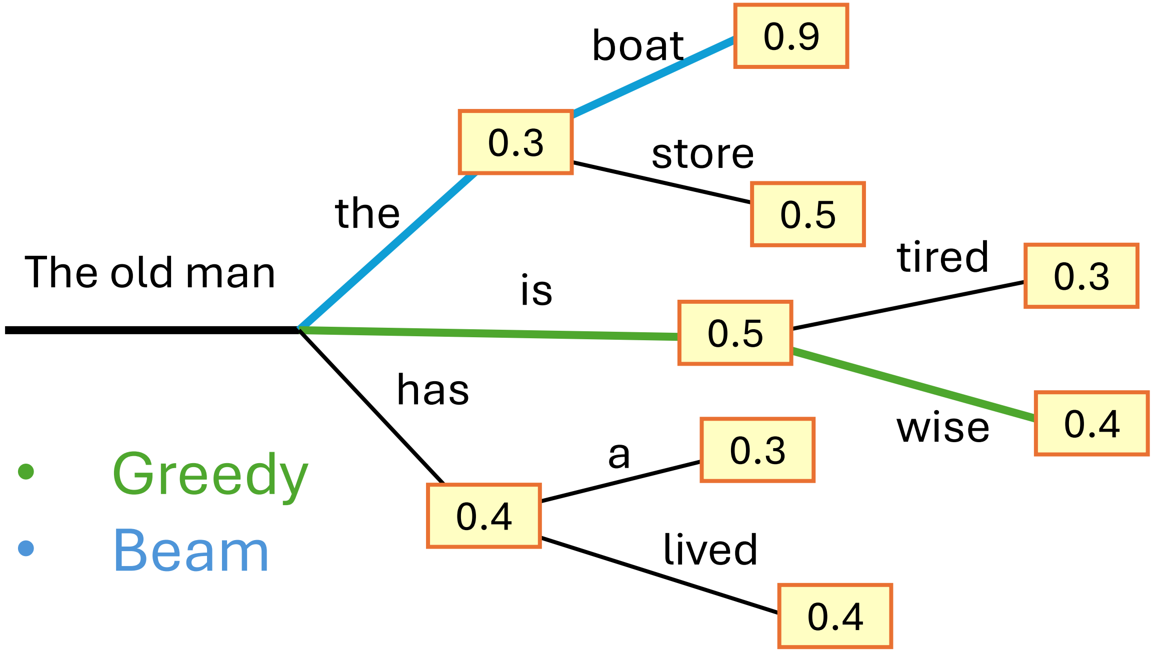
from **NLP**

“The old man the boat.”

Initially of less probable activities, which are redeemed by subsequent activities in the output sequence.

Solution Beam Search (1976)

At every step in the autoregression, a fixed number of best candidates is kept.



- Greedy
- Beam

Robustness Training

I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, “Generative adversarial nets”, NIPS14 (2014)

Adversarial methods

- **For next event:** F. Taymouri, M. La Rosa, S. M. Erfani, Z. D. Bozorgi, I. Verenich, “Predictive Business Process Monitoring via Generative Adversarial Nets: The Case of Next Event Prediction.”, BPM20 (2020).
- **For suffix:** F. Taymouri, M. La Rosa, S. M. Erfani, “A Deep Adversarial Model for Suffix and Remaining Time Prediction of Event Sequences.” SDM (2021).

Data augmentation

- **With noise insertion:** M. Käppel, S. Jablonski. “Model-Agnostic Event Log Augmentation for Predictive Process Monitoring”. CAiSE (2023)

Lesson learned?

~~We should definitely use this whole jumble of techniques together in a super complex model to aim for 99% accuracy!~~

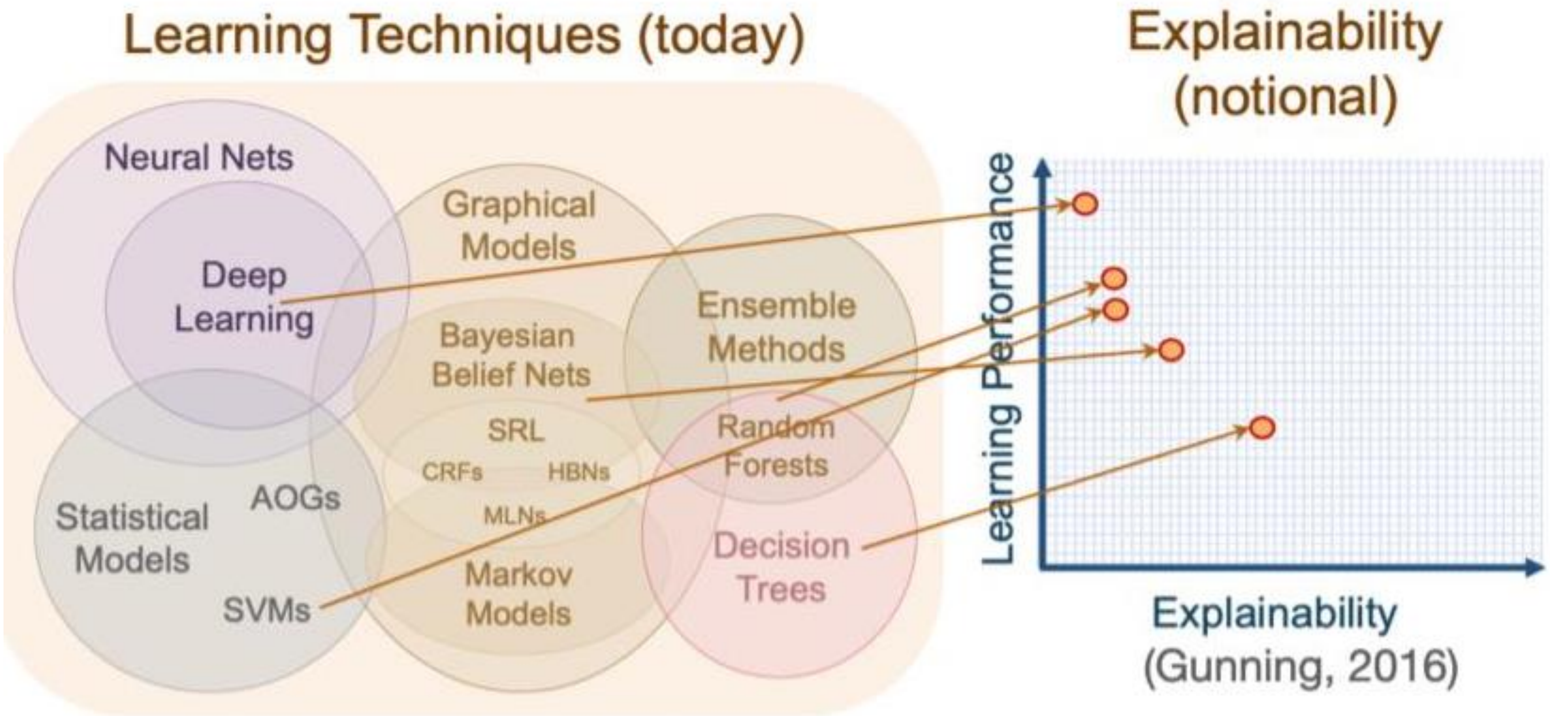
We want to look into the future!

How to Spot Dishonest Psychics

Visita ›

A final thought

Be careful not to create overly complex models for just a handful of accuracy points!



D. Gunning, D. Aha, “DARPA’s Explainable Artificial Intelligence (XAI) Program.” *AI Magazine (2019)*

Inter-case predictions

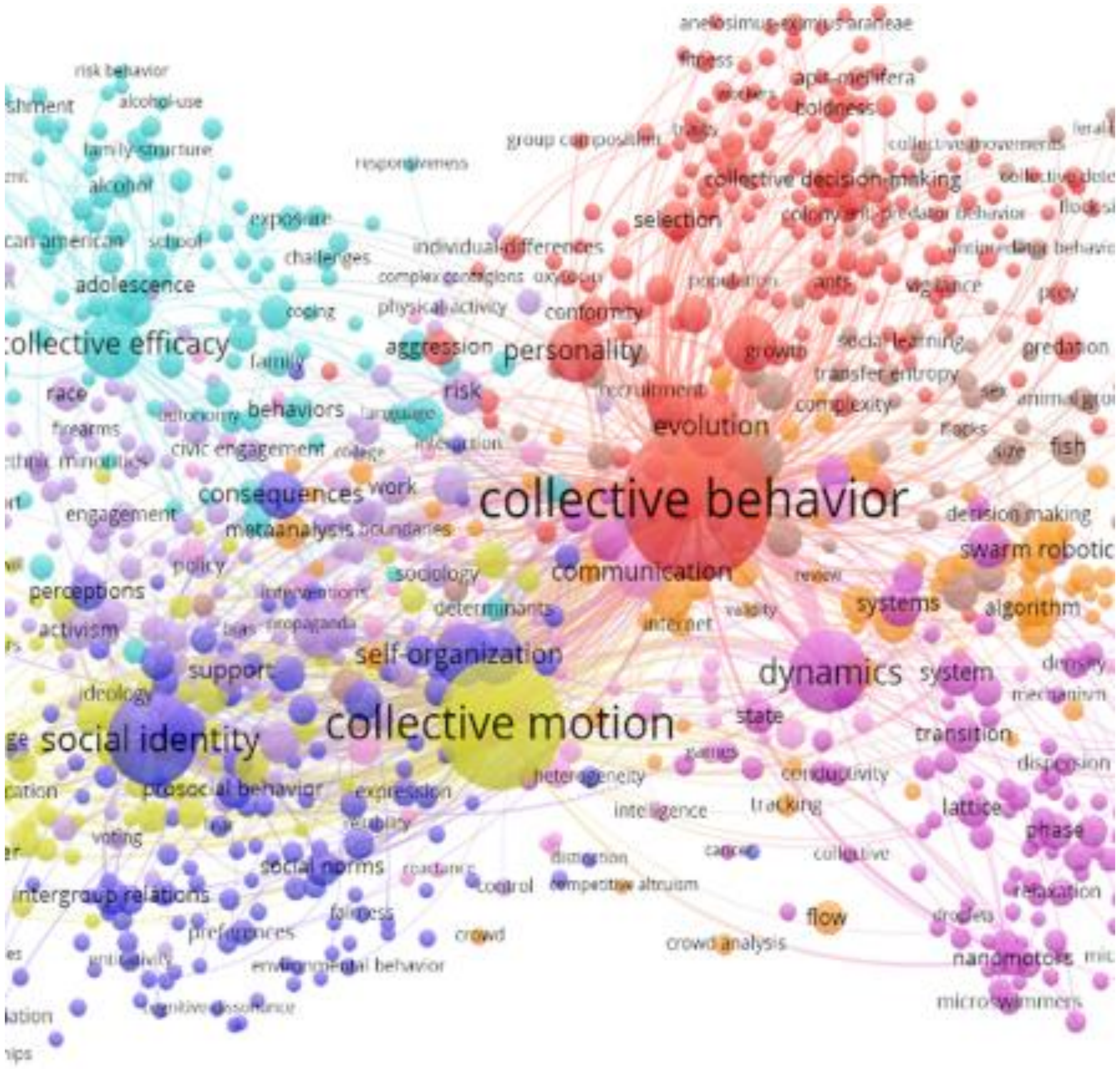
or, a trace is not always independent from other traces

- Traditional approaches make predictions taking into account a single execution.
- What if only limited resources are available?
- Predictions related to an ongoing case often also **depend on other cases (inter-case dependency)**.

Arik Senderovich, Chiara Di Francescomarino, Fabrizio Maria Maggi: **From knowledge-driven to data-driven inter-case feature encoding in predictive process monitoring**. Inf. Syst. 84: 255-264 (2019)

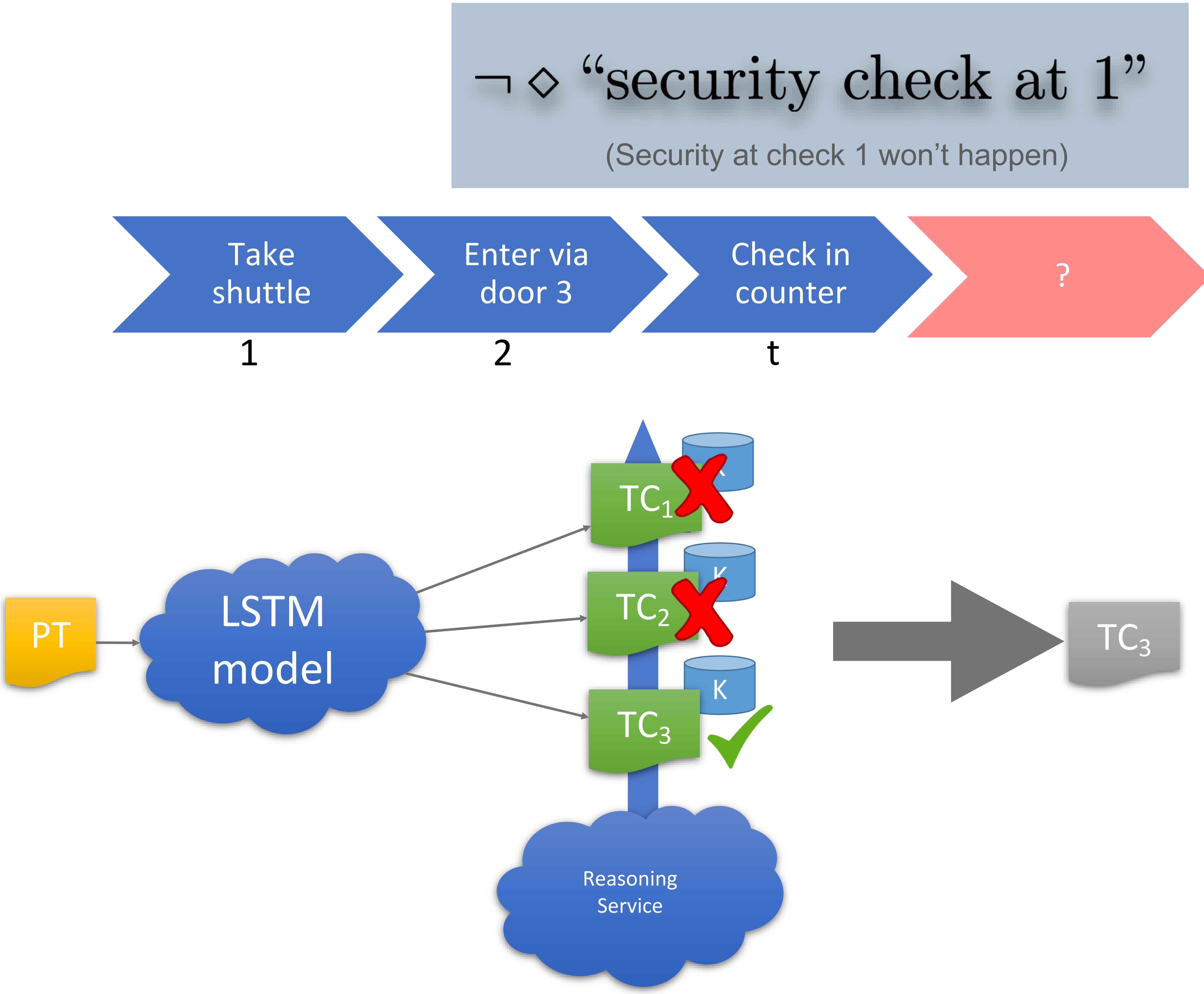
Collective behaviour

- Not many studies that investigate the system as a whole.



Predicting with hybrid architectures

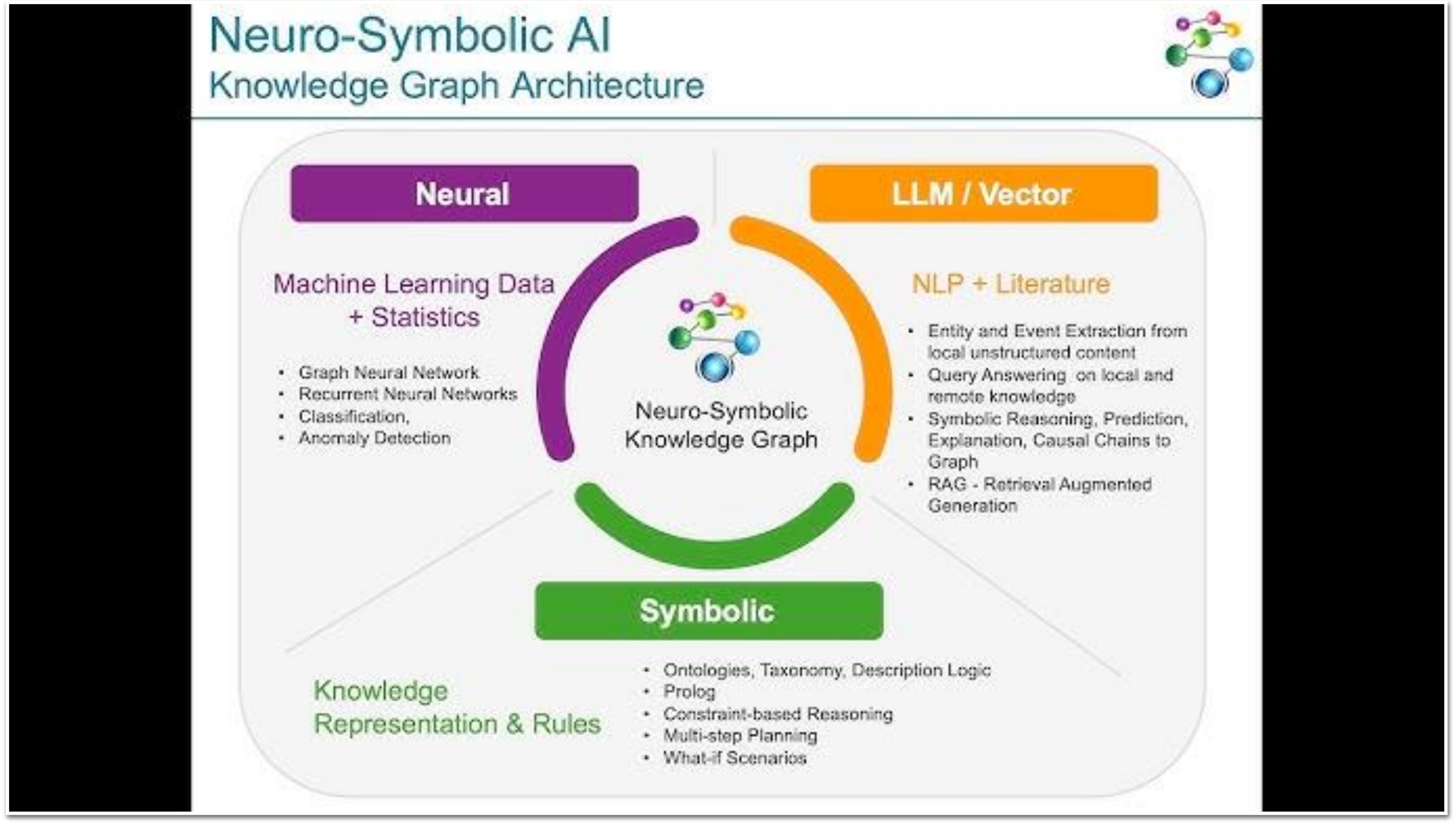
- Traditional approaches make predictions taking into account only ML models
- What if other ways of inference are available?
  - Reasoning
  - LLMs



Chiara Di Francescomarino, Chiara Ghidini, Fabrizio Maria Maggi, Giulio Petrucci, Anton Yeshchenko: **An Eye into the Future: Leveraging A-priori Knowledge in Predictive Business Process Monitoring**. BPM 2017: 252-268

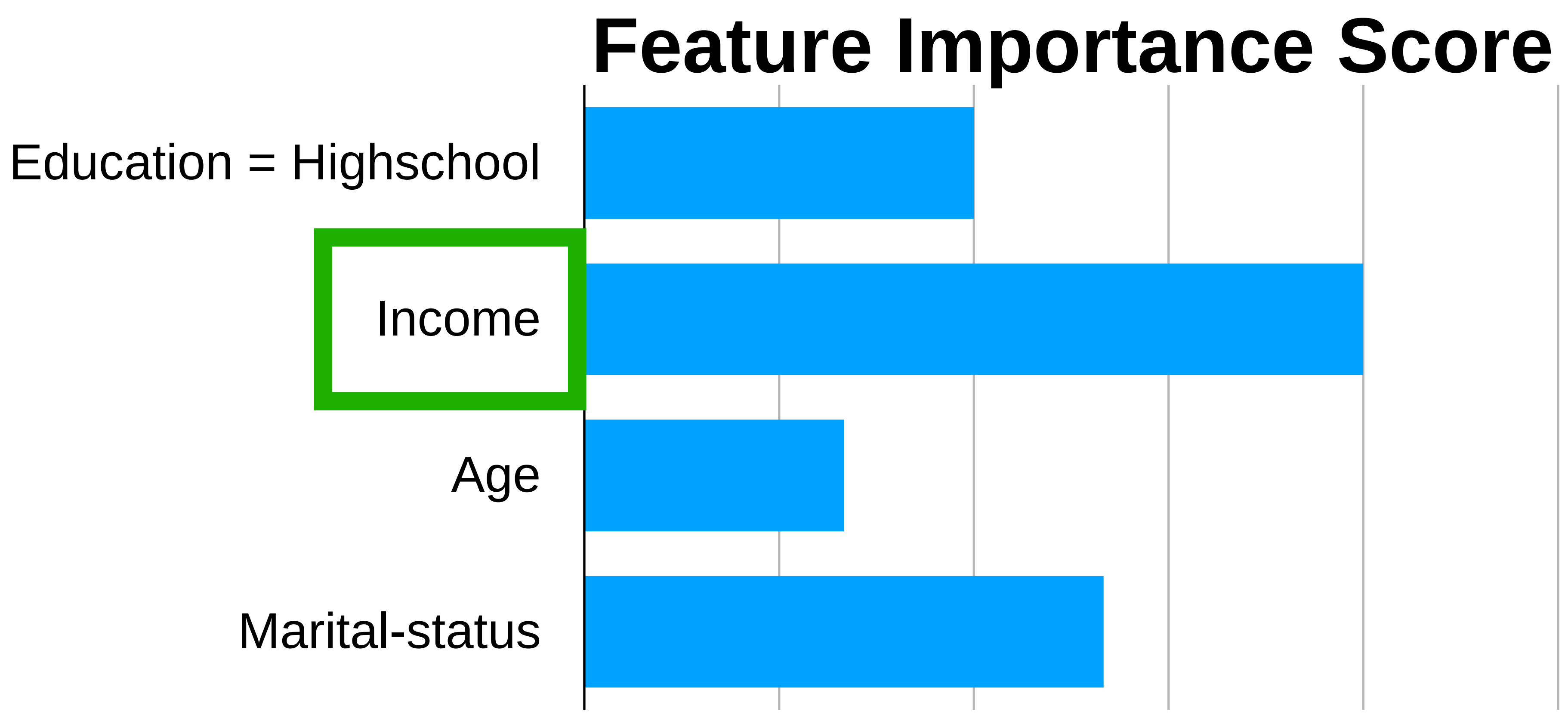
Neuro symbolic architectures

- No studies that investigate the development of neuro-symbolic architectures for our data.



Most work is on counterfactuals. Is this enough?

Feature importance techniques



Counterfactual explanations

Exploring "what-if" scenarios Watcher et al. (2017)

If your income was $5,000 higher, you would been granted the loan

Tool support

- Nirdizati
- Apromore
- ….

- Shall we have a repository of all our techniques?

Andrei Buliga, Riccardo Graziosi, Chiara Di Francescomarino, Chiara Ghidini, Fabrizio Maria Maggi, Williams Rizzi, Massimiliano Ronzani **Nirdizati Light: A Modular Framework for Explainable Predictive Process Monitoring** CEUR workshop proceedings

Thanks to

Chiara Di Francescomarino, Wil van der Aalst, Marlon Dumas, Marcello La Rosa, Anna Leontjeva, Fabrizio Maria Maggi, Williams Rizzi, Arik Senderovich, Irene Teinemaa, Ilya Verenich, Anton Yeshchenko, Marco Montali, Andrei Buliga, Massilmiliano Ronzani, ….