INESC TEC

INSTITUTO DE ENGENHARIA
DE SISTEMAS E COMPUTADORES,
TECNOLOGIA E CIÊNCIA

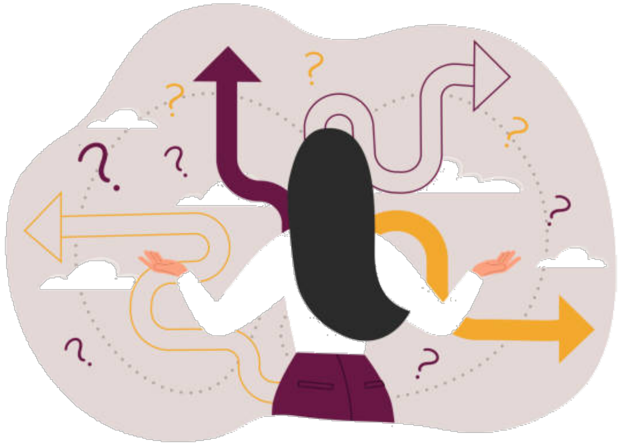
Towards interpretable models and human-AI integration in energy systems

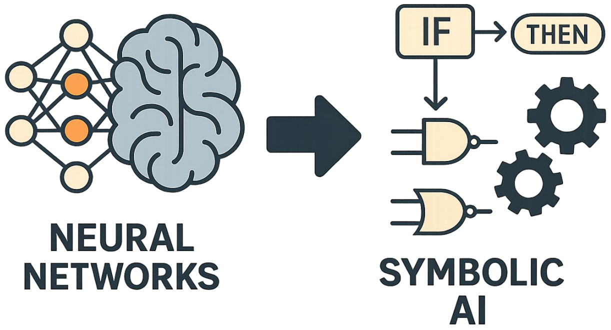
Ricardo Bessa

DTU PES Summer School 2026, Lyngby

18 May 2026

This lecture in a nutshell





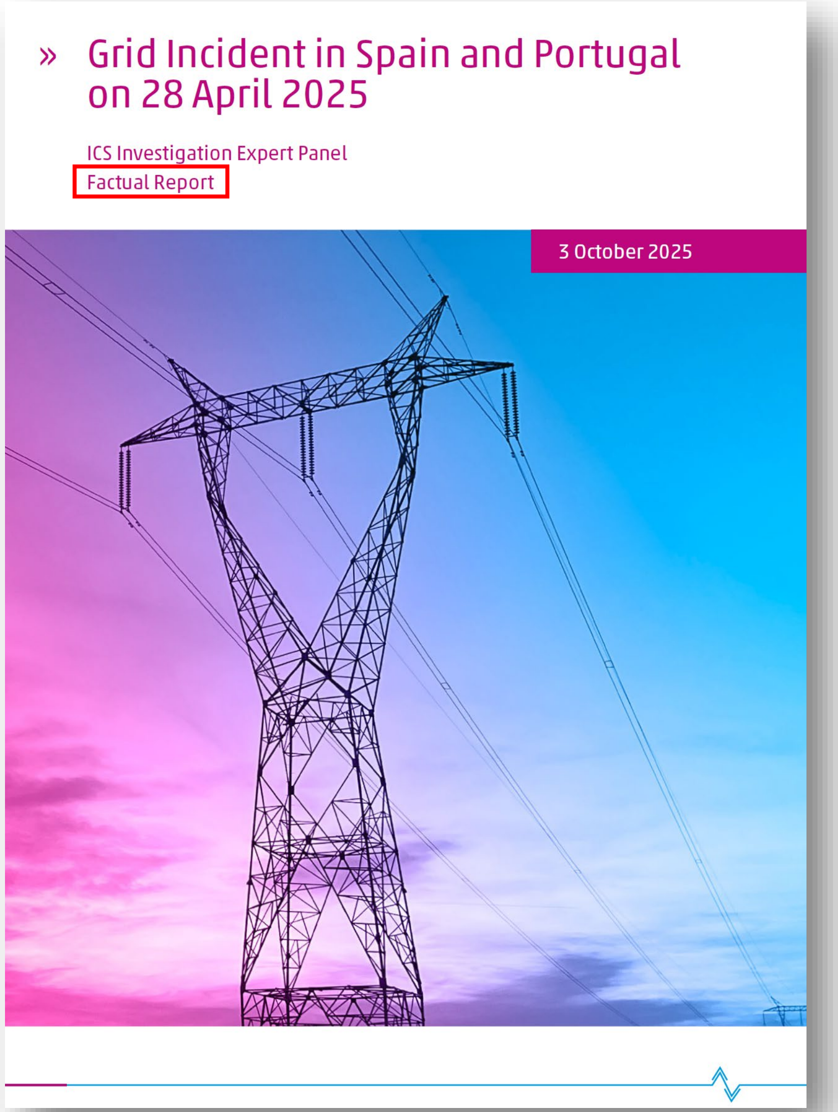
Human decision-making under forecast uncertainty

Neuro-symbolic learning for transparent data-driven decision-making

Application to energy systems

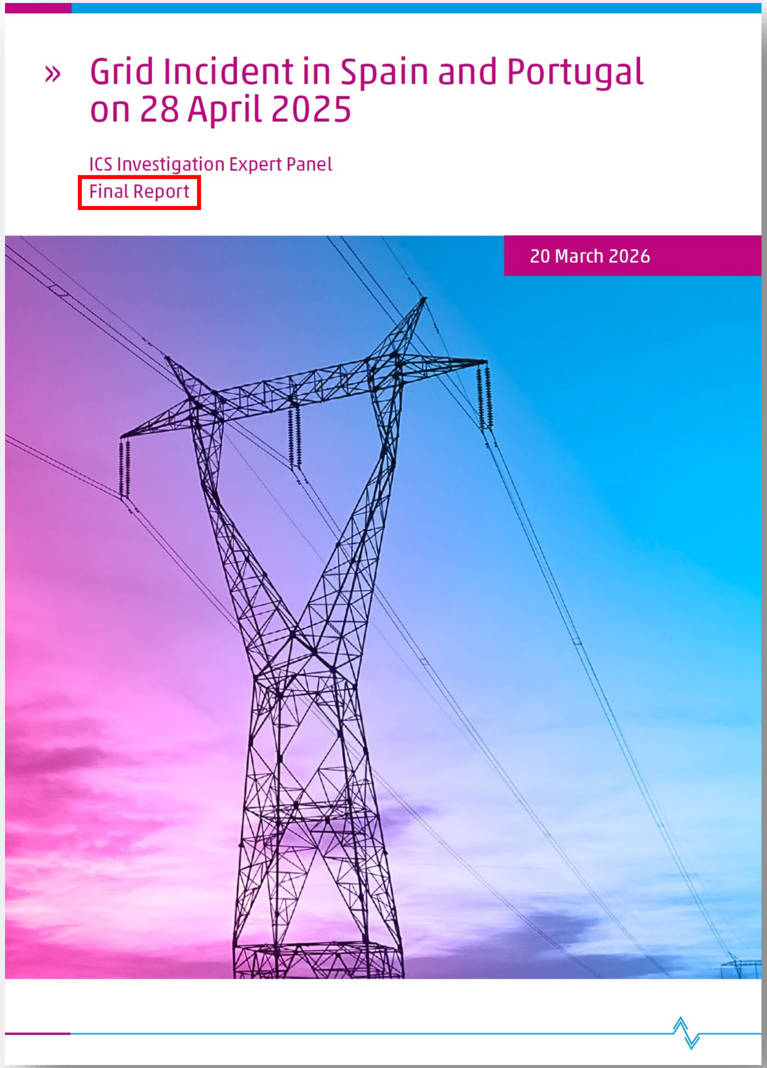
Human decision-making under forecast uncertainty

[Generated with ChatGPT based on Botterud, A., Wang, J., Miranda, V., Bessa, R. J. (2010). Wind power forecasting in US electricity markets. The Electricity Journal, 23(3), 71-82.]



~120 manual actions by the REE (TSO) in 210 min

"...the decision to perform the maneuvers is taken ***either based on the experience of the operator or based on a static power flow simulation*** *performed right before, to check the expected impact of the action on voltages and loads of elements."*



*“The manual corrective actions were carried out within **humanly realistic time frames**. The measured response times between the first recorded alarm exceeding the high-voltage alarm threshold and the execution of the **first voltage control remedial action 1min 33s in the northern zone and 2min 11s in the southern zone** take into account the time required for the operator to detect the alarm, analyse the issue, assess the most appropriate action while taking its impact into account, and implement it.”*

*“In the final analysed period, Red Eléctrica’s **dispatchers were unable to implement manual voltage control actions due to the very short duration of the event (38 sec)**. In such timeframes, only **dynamic voltage control** – i. e., generator voltage regulation together with STATCOM and HVDC voltage control – can effectively mitigate voltage deviations.”*

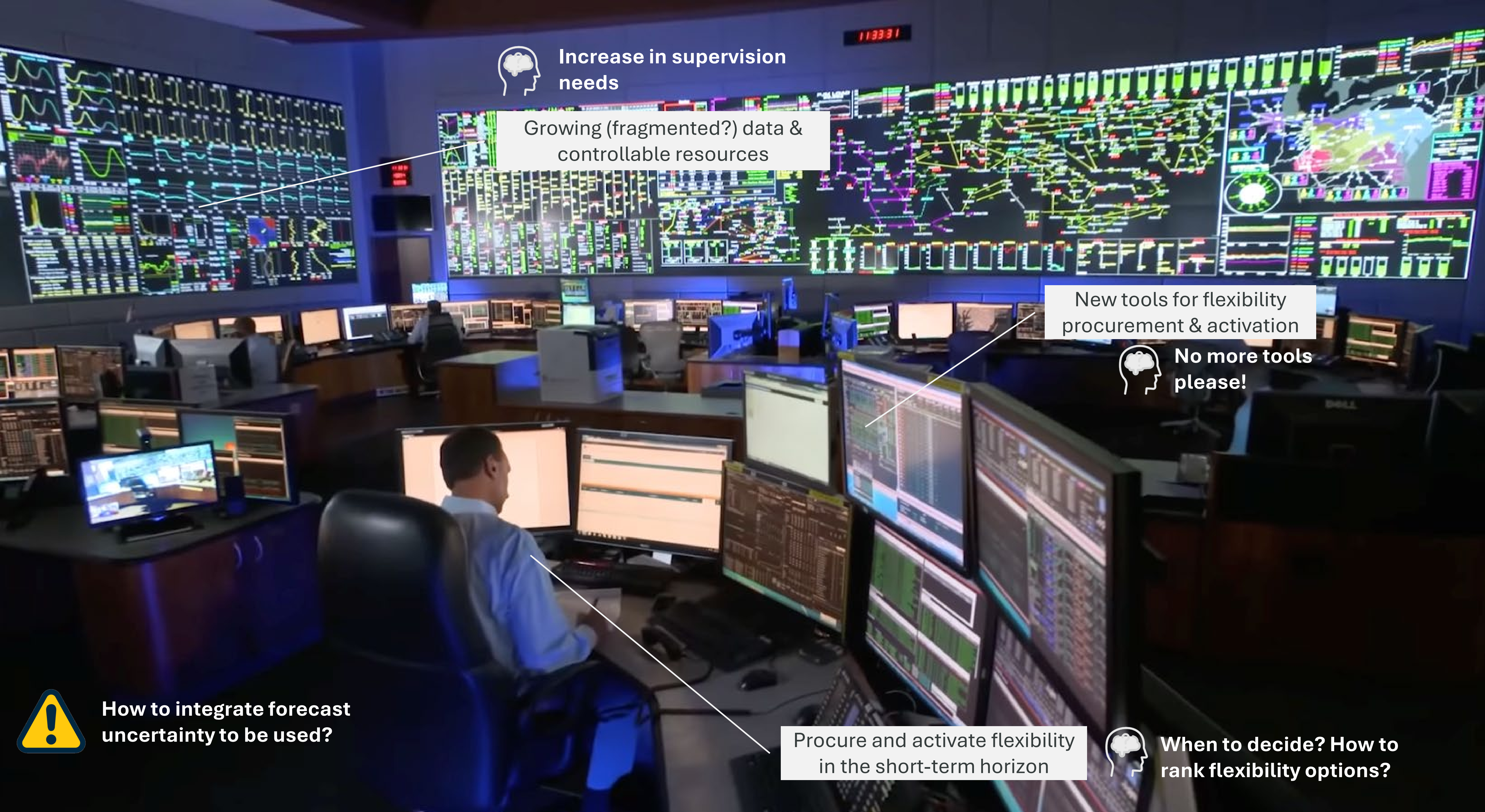
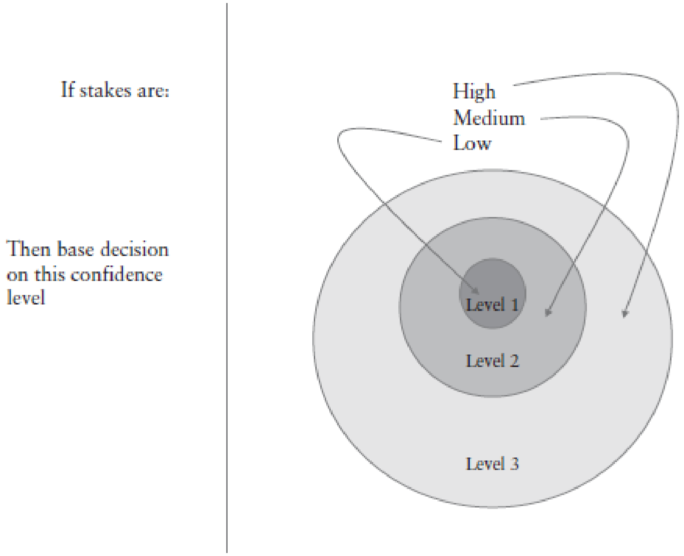


Image from: https://www.renewableenergyworld.com/power-grid/outage-management/building-the-control-room-of-the-future-to-ensure-resilient-power-grids/

Decision under forecast uncertainty: concepts

Confidence-based decisions

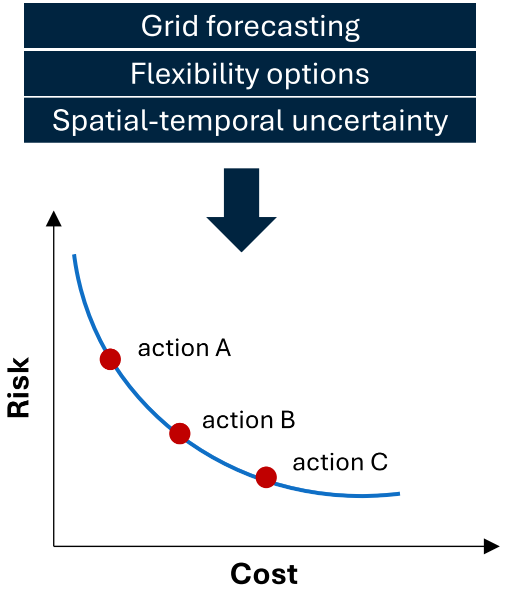


Source: Hill (2013). Confidence and decision. Games and Economic Behavior, 82, 675-692

Uncertainty forecasts with a larger spread can be helpful in catching low-probability-high-impact events, but can lead to expensive decisions due to high uncertainty

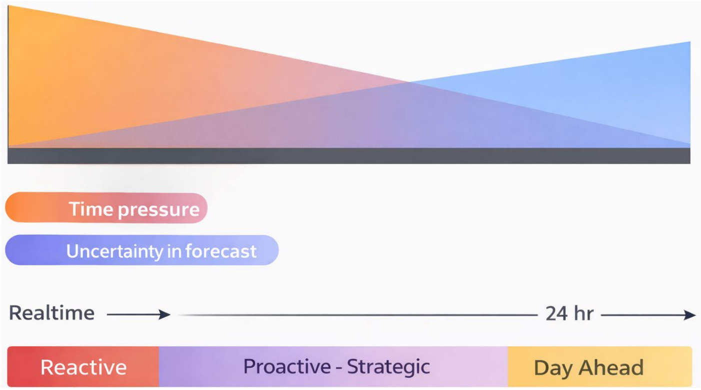
Narrow forecast intervals can, on the other hand, lead a decision-maker to overconfidence in a decision

Condensing large/complex information into risk curves



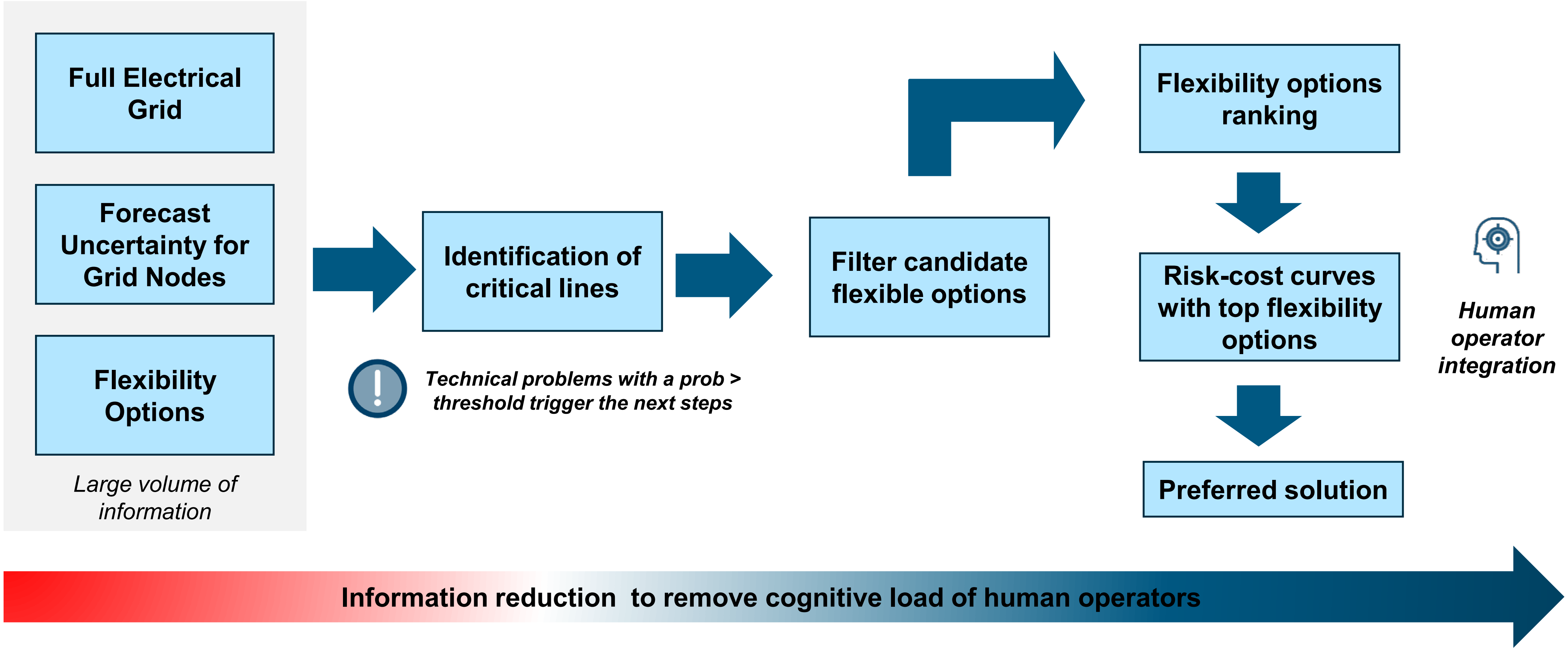
Meaningful information for human operators to decide

"Best" moment to decide



Key decision: wait (e.g., for a new forecast) or take an action now?

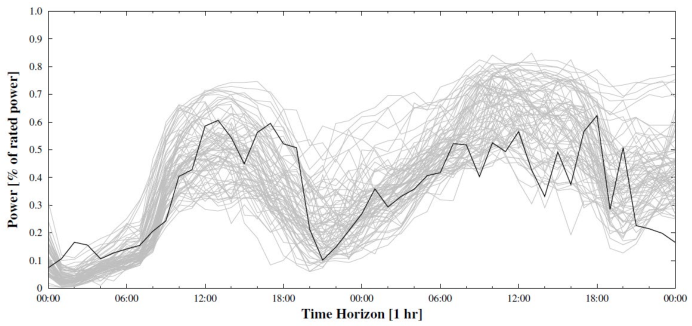
Framework: risk-aware flexibility procurement



Reference: R.J. Bessa, F. Moaidi, J. Viana, J.R. Andrade, "Uncertainty-aware procurement of flexibilities for electrical grid operational planning," *IEEE Transactions on Sustainable Energy*, vol. 15, no. 2, pp. 789-802, Apr. 2024

Building blocks: scenarios & sensitivities

Spatial scenarios generated with a Gaussian copula and marginal probability distributions



For each scenario, compute sensitivity indices relating

P, Q ~ Branch current (1)
P, Q ~ Node voltages (1)
NTW reconfiguration ~ Branch current (Z-bus + graph theory)

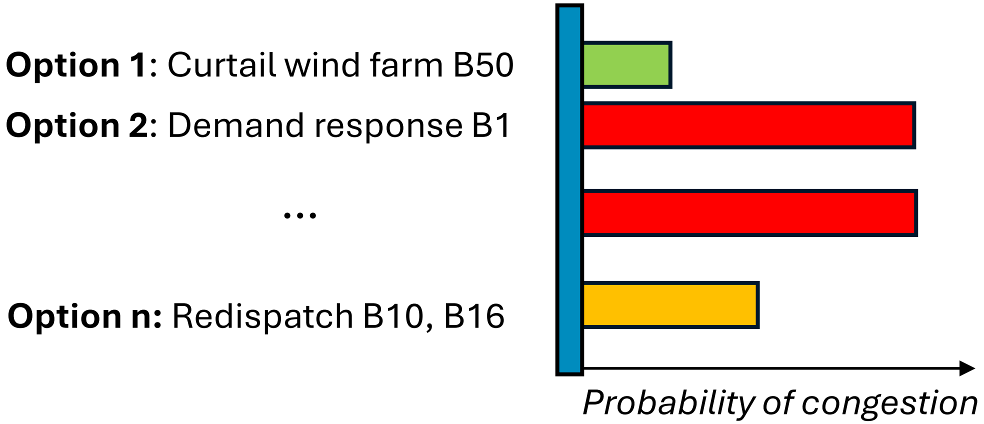
For each flexibility option, metrics are computed *to characterize its* ***effectiveness***

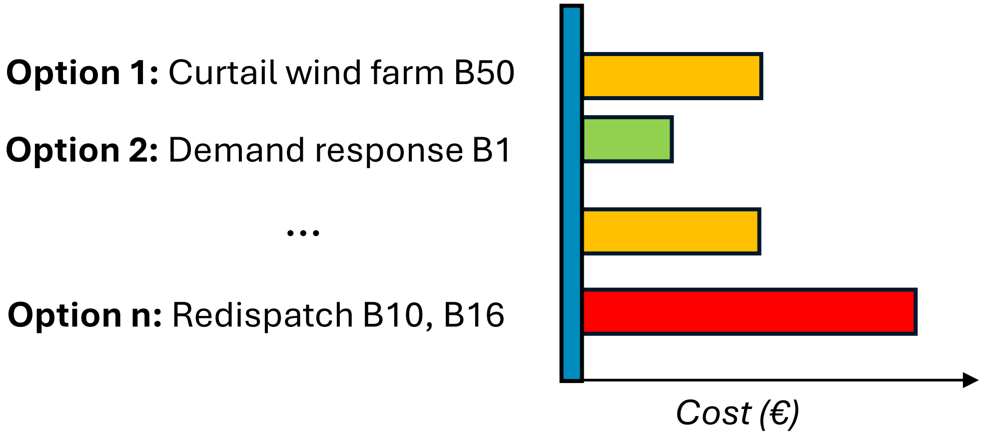
- **Expected flexibility cost**
- **Probability of congestion**
- **VaR of flexibility cost**
- **VaR of severity**
- **Expected severity**

(1) K. Christakou, et al. "Efficient computation of sensitivity coefficients of node voltages and line currents in unbalanced radial electrical distribution networks," *IEEE Trans. on Smart Grid*, vol. 4(2), pp. 741-750, June 2013

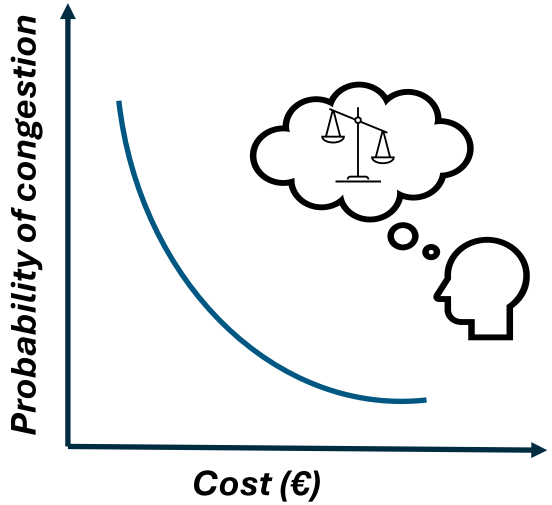
Building blocks: flexibility ranking, risk curves

Flexibility options ranking with TOPSIS[1]





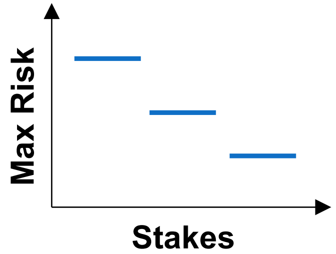
Combine top *k* flexibility options



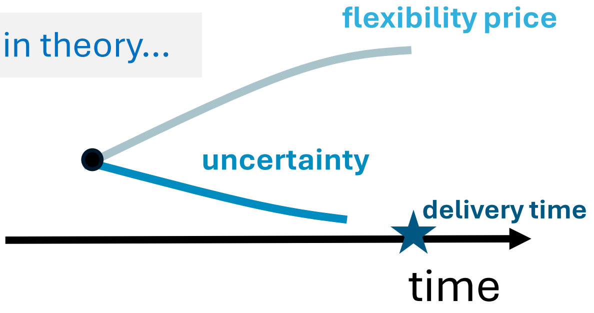
(1) K. Yoon, "A reconciliation among discrete compromise situations," *J. of the Op. Res. Soc.*, vol. 38, pp. 277–286, March 1987.

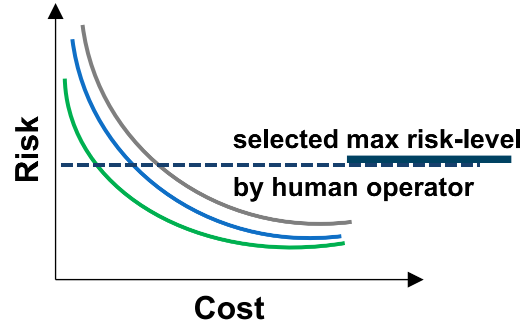
Time-to-decide (T2D) and meta-forecasts

Probability of a congestion forecasted with NWP for day D+1 (lead time: t+30)
> Decide now (“reserve” a flexibility option) or wait for next forecast?

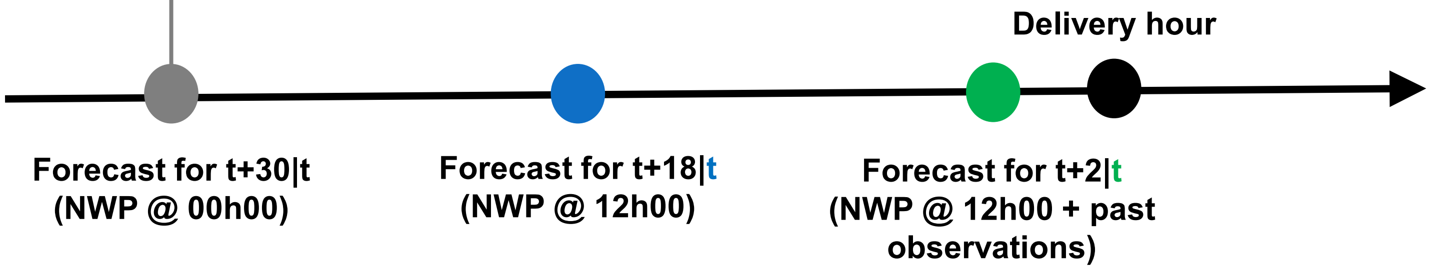


in theory...



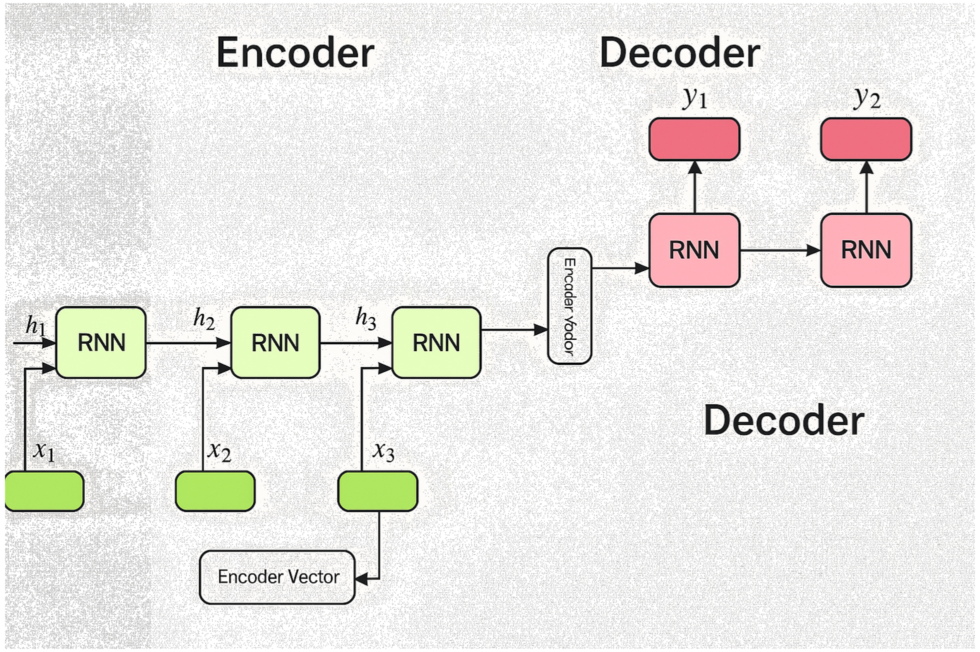


At decision time *t*, generates
- ❑ Forecast for t+30|t
- ❑ Meta-forecast for t+30|t+12
- ❑ Meta-forecast for t+30|t+28



Meta-forecasting model

Forecasted generated with 00h00 NWP
+
Features characterizing level uncertainty (IQR, forecasted quantiles, stdev.)

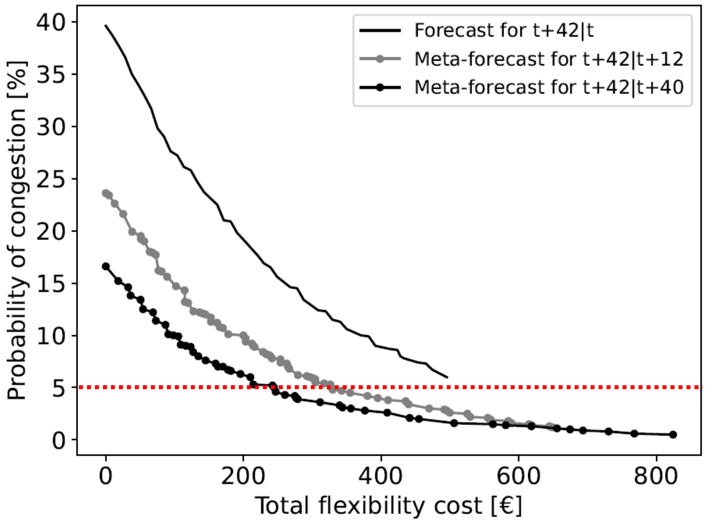


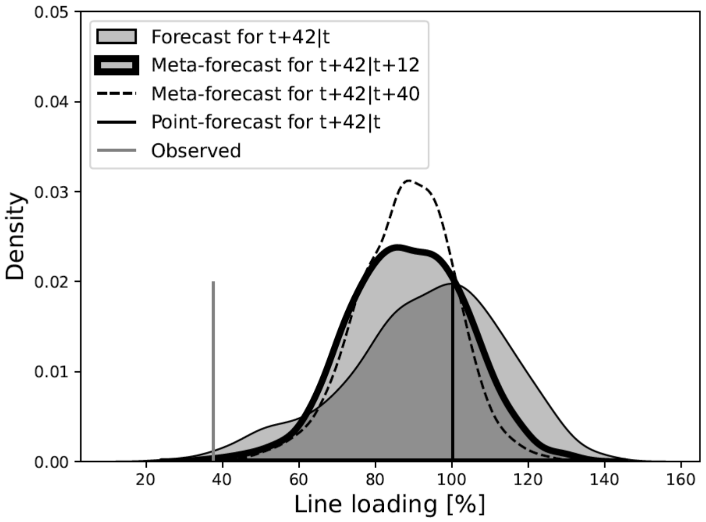
baseline** model: **forecast does not change

- ❑ MAE improvement (**meta-forecast with NWP @ 12h00**) between 13% and 26%
- ❑ MAE improvement (**meta-forecast for t+2|t**) between 16% and 31%

Flexibility management use case example

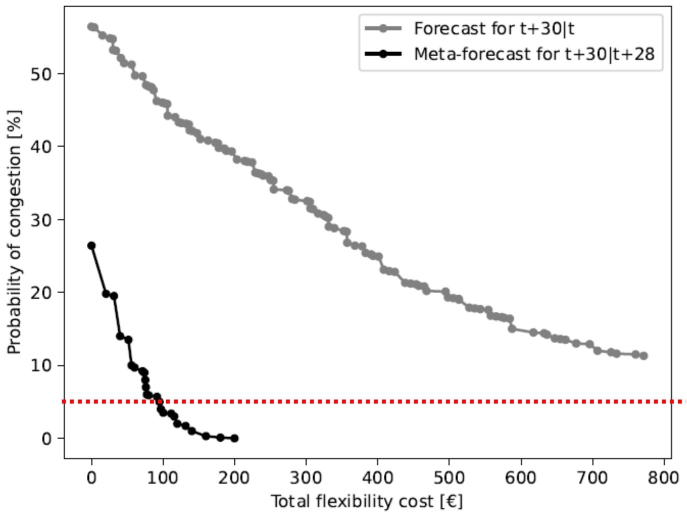
Forecast and meta-forecast launched at 00h00 for t+42|t





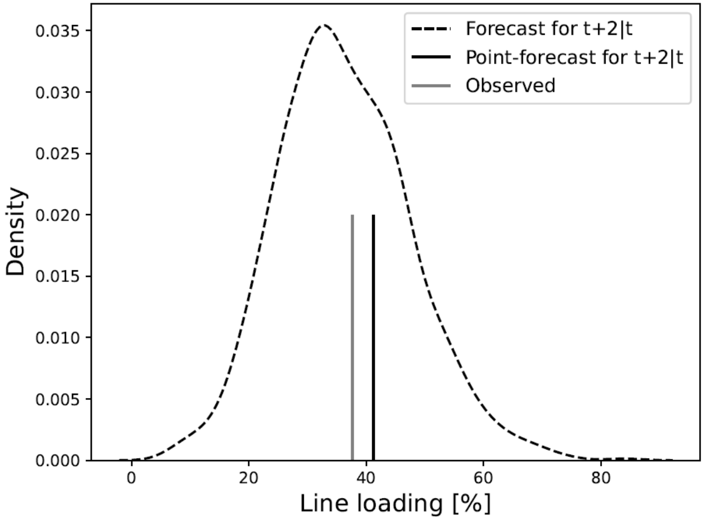
Wait for the next forecast update. DO NOTHING

forecast and meta-forecast launched at 12h00



Wait for the next forecast update. DO NOTHING

forecast launched for t+2|t



No congestion (no need for flexibility) Saved flexibility cost!

INESC TEC

Results for Oberrhein MV network

KEY MESSAGES

- ❑ **A mix of physics-based information, data and interaction with the decision-maker offers a simple approach to rank and select flexibility options** *(e.g., example from industry is TenneT GridOptions)*
- ❑ **Time-to-decide (T2D) approach outperforms a decision-now strategy** (*i.e., operator decides to reserve flexibility at the lowest price*)
  - Improves by 30% the cost-loss matrix performance metric

Cost-loss matrix

	Congestion occurred	Congestion did not occur
Action taken	Rate of occurrence in % (h) **Flex cost (C) + Loss (L)**	Rate of occurrence in % (m) **Flex cost (C)**
Action not taken	Rate of occurrence in % (f) **Loss (L)**	Rate of occurrence in % (c) **No cost**

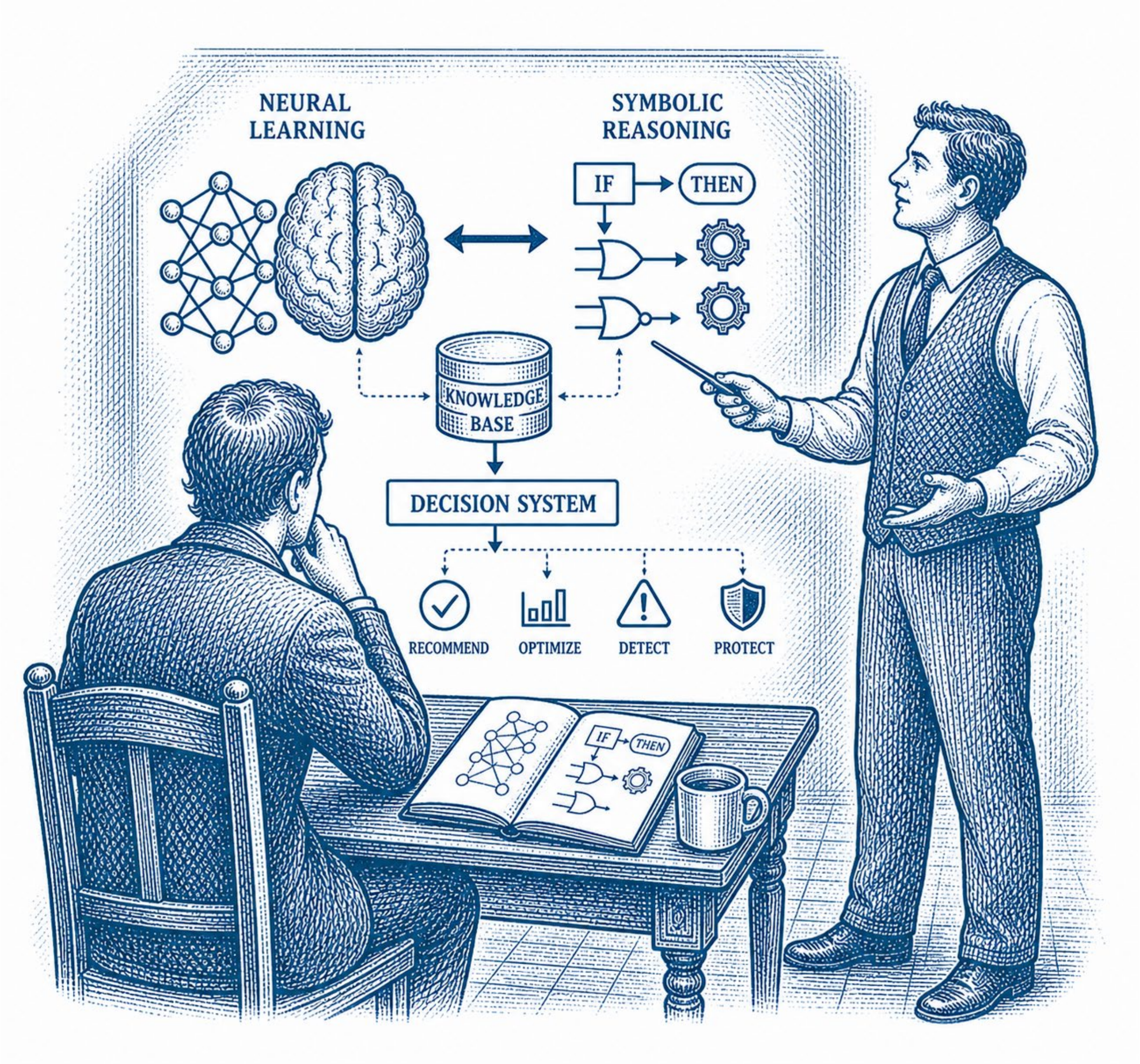
Metric

$$\gamma = (C + L) \cdot h + C \cdot m + L \cdot f + 0 \cdot c$$

This is **just combining information** in a step-by-step approach and not evolving knowledge from data

A more advanced learning approach is needed...

Neurosymbolic learning for interpretable data-driven decision making



[generated with ChatGPT]

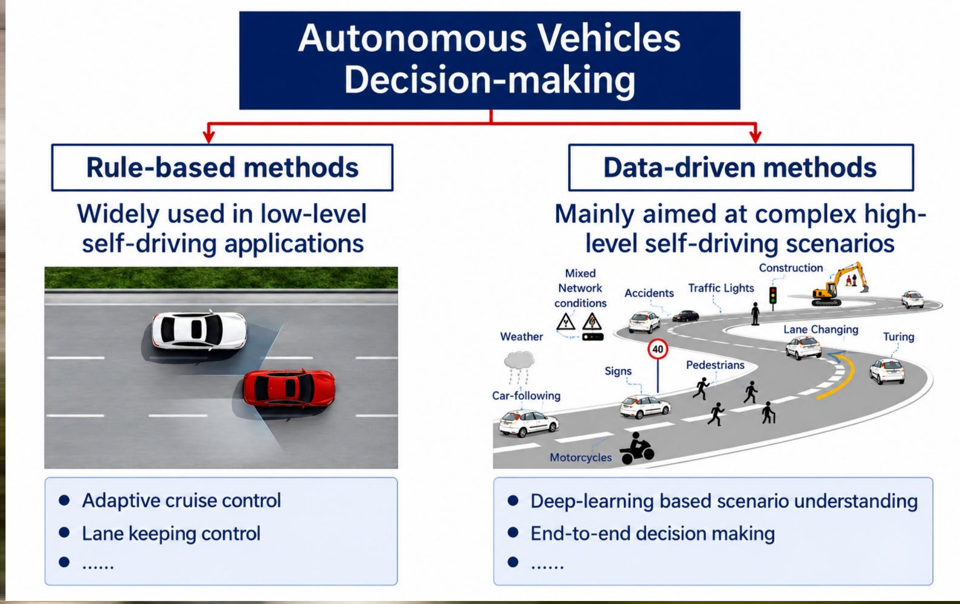
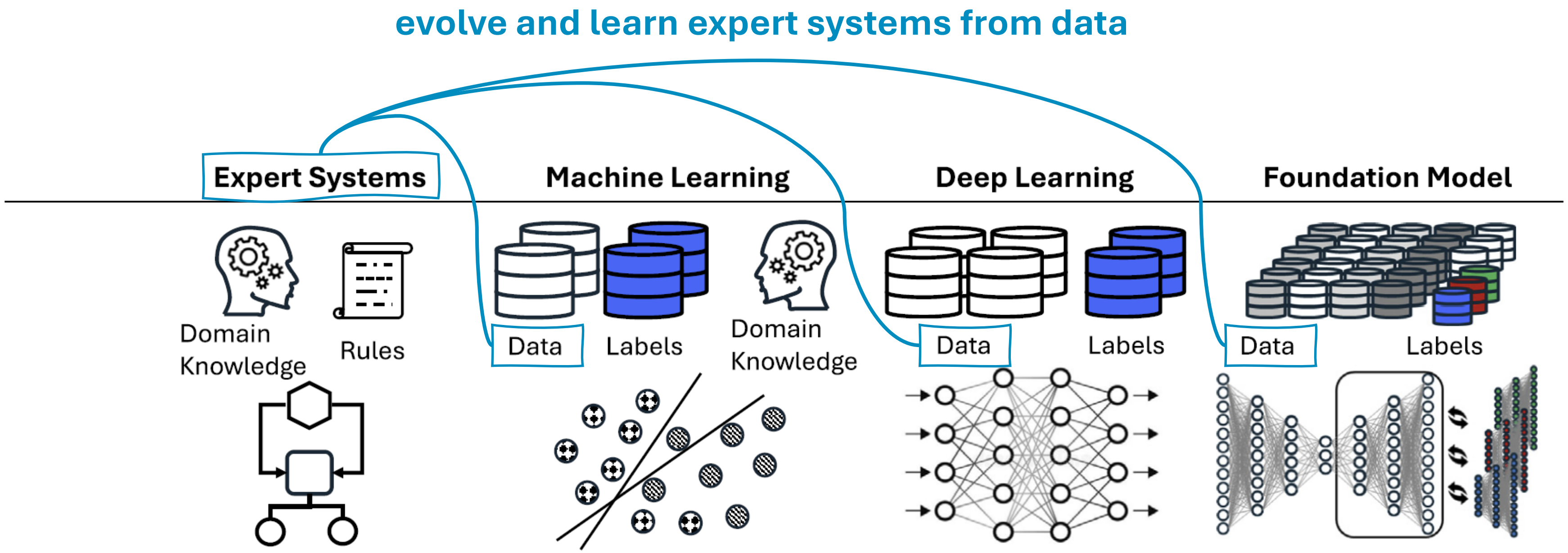


Image from: https://arxiv.org/abs/2306.16784

Image from: https://www.wsj.com/tech/this-time-the-hype-around-self-driving-cars-feels-real-e37f2e7f

Opportunity along the AI evolution chain

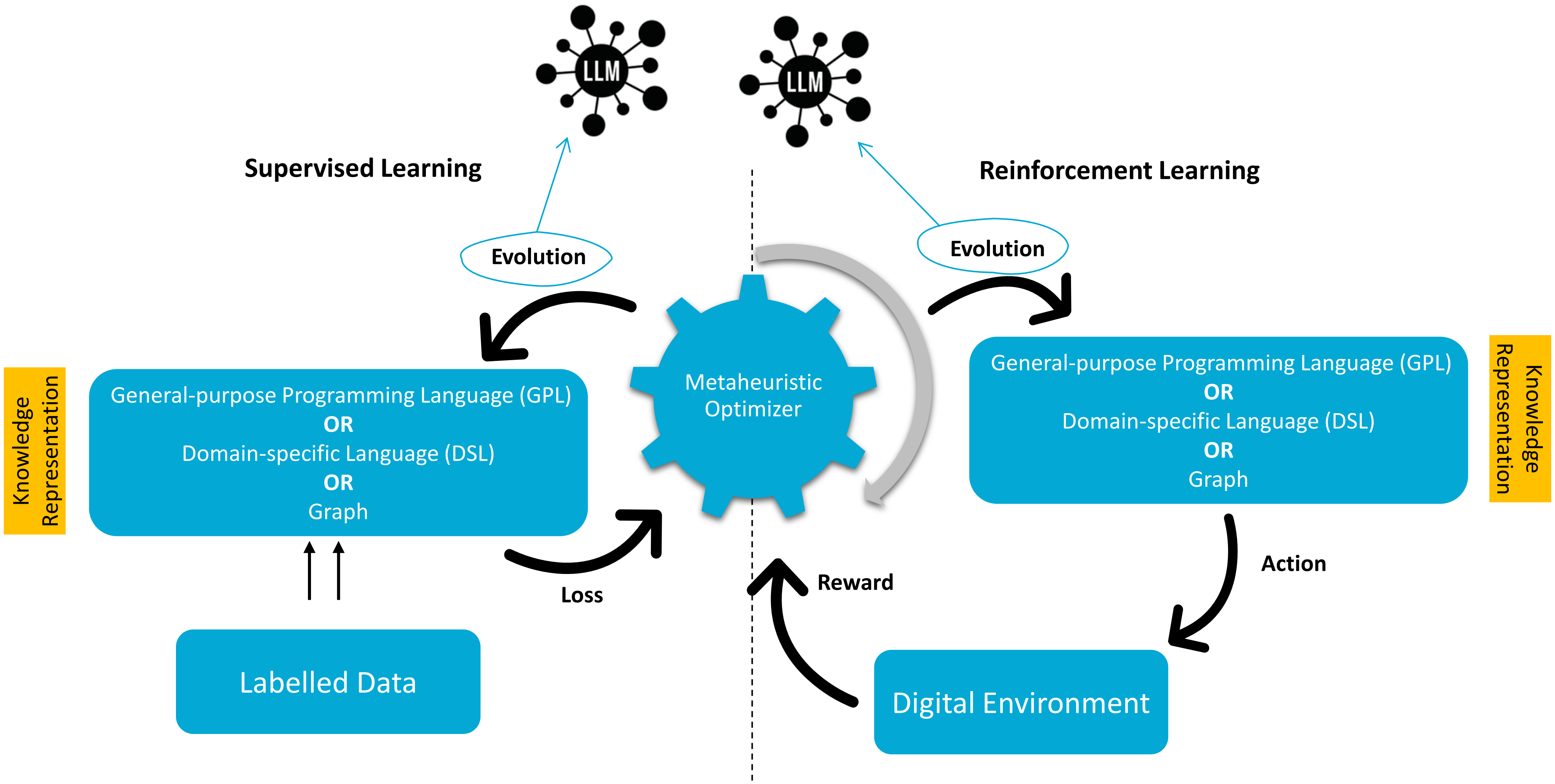


	Expert Systems	Machine Learning	Deep Learning	Foundation Model
Model Size	Very small	Small	Large	Very Large
Features	Hand-designed	Hand-designed	Learnt	Learnt
Learning	None	Supervised	Supervised	Self-supervised
Data	Very few	Few labeled Data	Labeled data	Very large unlabeled + small labeled data
Adaptability	None	Little	Medium	Large

Figure from: H.F. Hamann, et al., "A perspective on foundation models for the electric power grid," *Joule*, vol. 8(12), pp. 3245-3258, Dec. 2024.

It's no longer just about learning from data, but about the **evolution of knowledge from it**

Neurosymbolic learning



Reference: F. Fernandes, R.J. Bessa, J.A. Peças Lopes, "Evolving symbolic model for dynamic security assessment in power systems," *Journal of Modern Power Systems and Clean Energy*, vol. 13, no. 4, pp. 1113-1126, 2025.

(1) Application to dynamic security assessment

Image from: https://news.cision.com/abb-power-grids/i/control-room,c2661365

General-purpose Programming Language (GPL)

Reference: F. Fernandes, R.J. Bessa, J.A. Peças Lopes, "Evolving symbolic model for dynamic security assessment in power systems," *Journal of Modern Power Systems and Clean Energy*, vol. 13, no. 4, pp. 1113-1126, 2025.

Evolving symbolic model (ESM) for dynamic security assessment

template evolution process using **Simulated Annealing**

```
  if { Conditional Statements } :
       { Sequence of Action }
elif { Conditional Statements }:
       { Sequence of Action }
else:  { Sequence of Action }
```

Example of a simplistic template for ESM

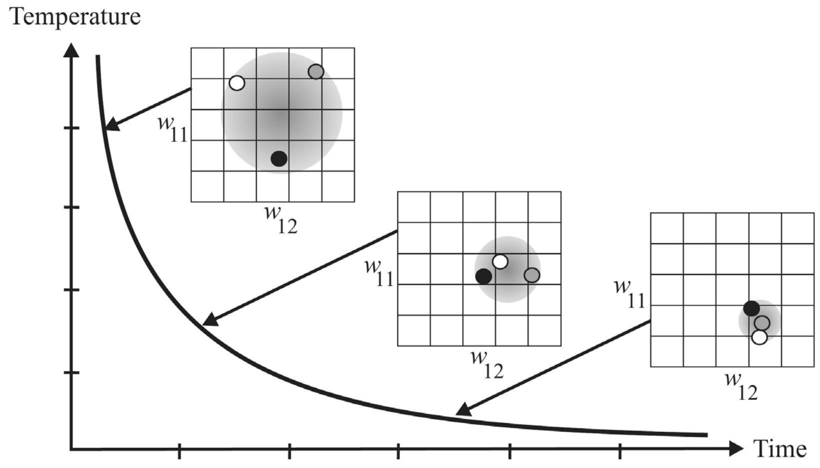
$$S_E \cup \{V_E, N_E, A_E, C_E, L_E, W_E\}$$
$$V_E = \{a, b, c, d, x, y, z\}\,,\ N_E = \{\mathbb{R}\}$$
$$A_E = \{+, -, \times, \div, =\}\,,\ C_E = \{<, >\}$$
$$L_E = \{and, or\}\,,\ W_E = \{if, elif, else\}$$

Example of elements to create new models

Algorithm 1 Evolution process of the ESM

Require: Kb, T_0, T_f

1: Generate a symbolic model, SM_i (II-B).
2: Tune numerical constants of SM_i (DE).
3: **while** $T > T_f$ **do**
4: Generate neighbor symbolic model, SM_n (III-B).
5: Tune numerical constants of SM_n (DE).
6: **if** $\mathcal{F}_i > \mathcal{F}_n$ **or** $\mathcal{R}(0,1) < e^{\frac{\mathcal{F}_i - \mathcal{F}_n}{T \cdot Kb}}$ **then**
7: SM_i=SM_n
8: Adjust T according to cooling scheme
9: Perform simplification on SM_i (III-C).



Source: https://www.intechopen.com/chapters/38520

ESM for dynamic security assessment

INESC TEC

Mutation	P_a
Addition/Removal of Condition *e.g.,* $if(x+y>1) \rightarrow if(x+y>4$ **AND** $z<3)$	1/15
Logical Operator Change *e.g.,* $if(x+y>4$ AND $z<3) \rightarrow if(x+y>4$ **OR z** $<$ **3**$)$	1/10
Comparison Operator Change *e.g.,* $if(x+y>4) \rightarrow if(x+y<4)$	1/5
Addition or Removal of Parcels *e.g.,* $if(x+y>4) \rightarrow if(x+y-\mathbf{z}>4)$	1/3
Parcel Mutation *e.g.,* $if(x>4) \rightarrow if(\mathbf{z}>4) \mid if(\mathbf{x}\times\mathbf{z}>4) \rightarrow if(\mathbf{x}\div\mathbf{z}>4)$	1

Algorithm 1 Evolution process of the ESM

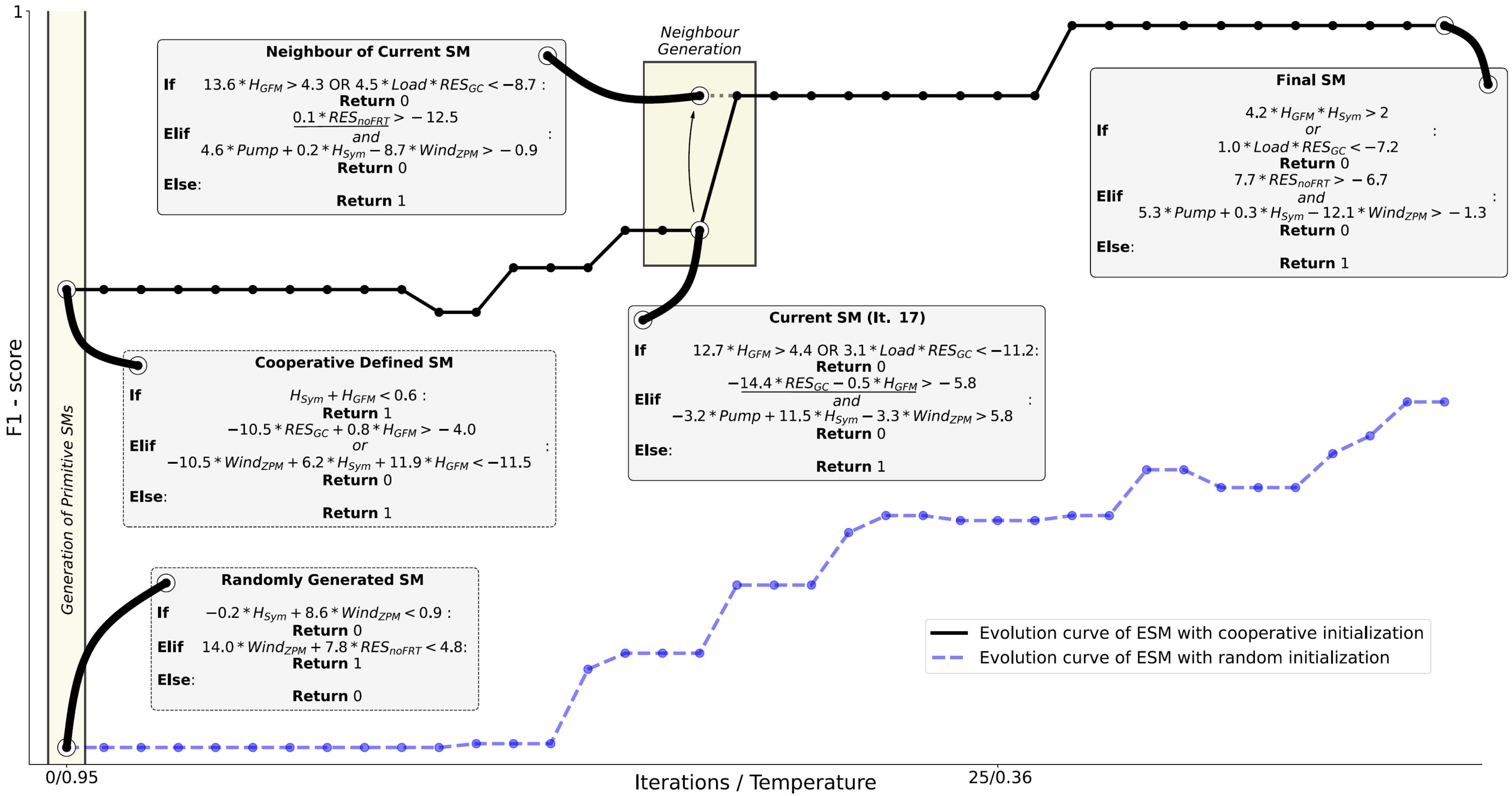
Require: Kb, T_0, T_f

Generate a symbolic model, SM_i (II-B).
Tune numerical constants of SM_i (DE).
while $T > T_f$ **do**
    Generate neighbor symbolic model, SM_n (III-B).
    Tune numerical constants of SM_n (DE).
    if $\mathcal{F}_i > \mathcal{F}_n$ **or** $\mathcal{R}(0,1) < e^{\frac{\mathcal{F}_i - \mathcal{F}_n}{T \cdot Kb}}$ **then**
        SM_i=SM_n
    Adjust T according to cooling scheme
    Perform simplification on SM_i (III-C).

Algorithm 2 *Simplification* process of the ESM

Output based simplification:
for All Statements (St_i) **do**
    Evaluate the effect of St_i on the SM output.
    if St_i does not affect the SM output **then**
        Remove part or the totality of statement St_i.
Mathmatical Inspection:
for All Statements (St_i) **do**
    Remove model parcels with multiplications by 0.
    `e.g., if(`$3\times x+0\times y$`>1): → if(`$3\times x$`>1):`
    Sum all zero-degree polynomials in each expression.
    `e.g., if(`$x+y-3$`>1): → if(`$x+y$`>4):`

ESM classification model: secure or insecure?



ESM decision model: preventive action

(trade-off between the cost of additional synchronous inertia and the enhanced system security)

$$R = \sum_{i=0}^{N_T} R_i = \sum_{i=0}^{N_T} (1 - 0.25(S_i + 1)^2 - K_t \times H_{sa_i})$$

reward function

Primitive model

if (S>0):
  if { $H_{GFM} < 1.92$ }:
    $\mathbf{H_{sa}}$ = { Action expression }
    $\mathbf{H_{va}}$ = { Action expression }
  else:
    $\mathbf{H_{sa}}$ = { Action expression }
    $\mathbf{H_{va}}$ = { Action expression }

Decision system

if (S>0):
  if $(1.6 \times (S \times H_{Sym}) + 0.33 \times (H_{Sym}/Res_{IP})) < 1.42)$:
    $\mathbf{H_{sa}} = 1.32 \times (PV\_ZPM/H_{Sym})$
    $\mathbf{H_{va}} = 2.02 \times (Load/Res_{IP})$
  else:
    $\mathbf{H_{va}} = 13.11 \times S$

Case study: Madeira island


14 novembro 2012 às 17h18

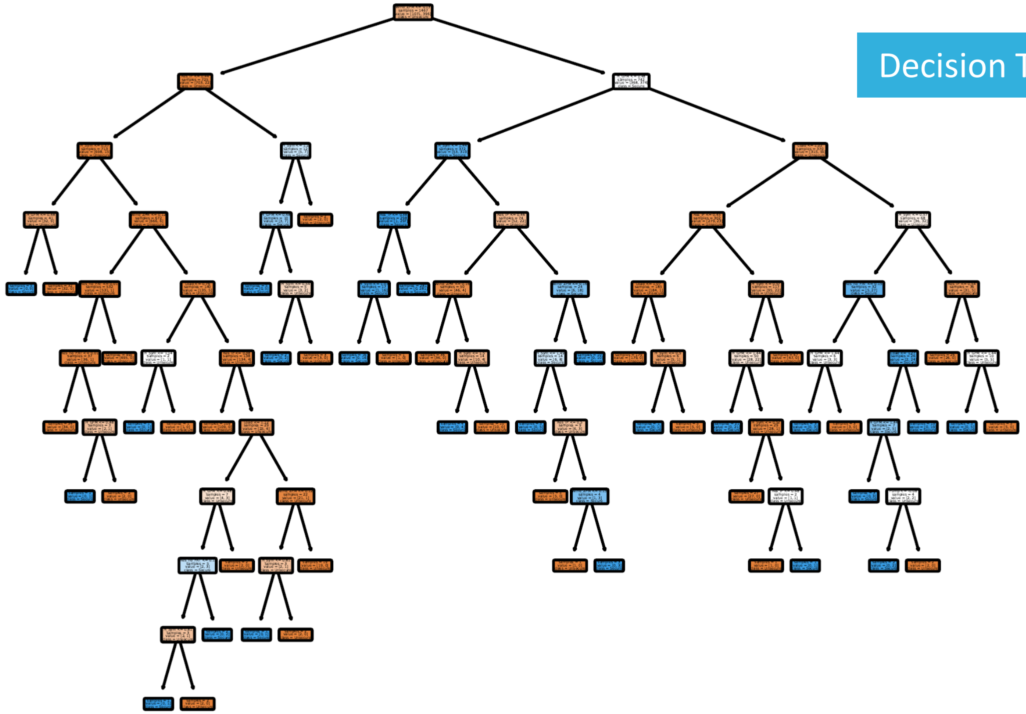
"Apagão" deixa ilha da Madeira sem eletricidade

Uma avaria numa máquina de uma das centrais elétricas da Madeira deixou a ilha sem eletricidade hoje à tarde. Segundo o DN apurou, a prioridade é restabelecer o fornecimento a todos os clientes de forma faseada.

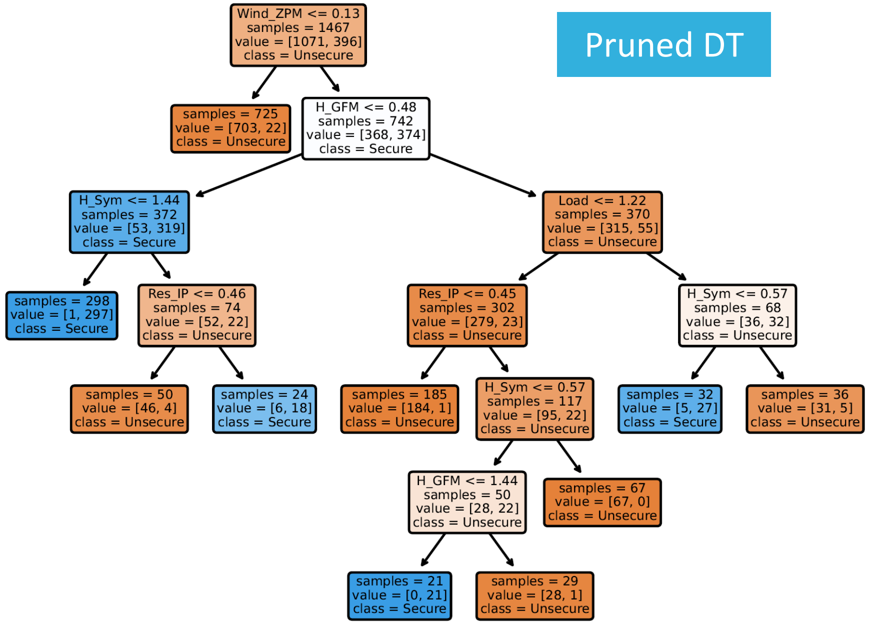
Diário de Notícias

Results: Interpretability

Decision Tree (DT)



Pruned DT



Evolving symbolic model (ESM)

if $\{0.7 \times \frac{H_{GFM}}{Wind_{ZPM}} + 4.03 \times \frac{H_{Sym}}{Load} < 6.40 \;\; AND \;\; 3.56 \times Res_{noFRT} \times H_{Sym} - 5.87 \times Res_{IP} < -2\}$:

return {1}

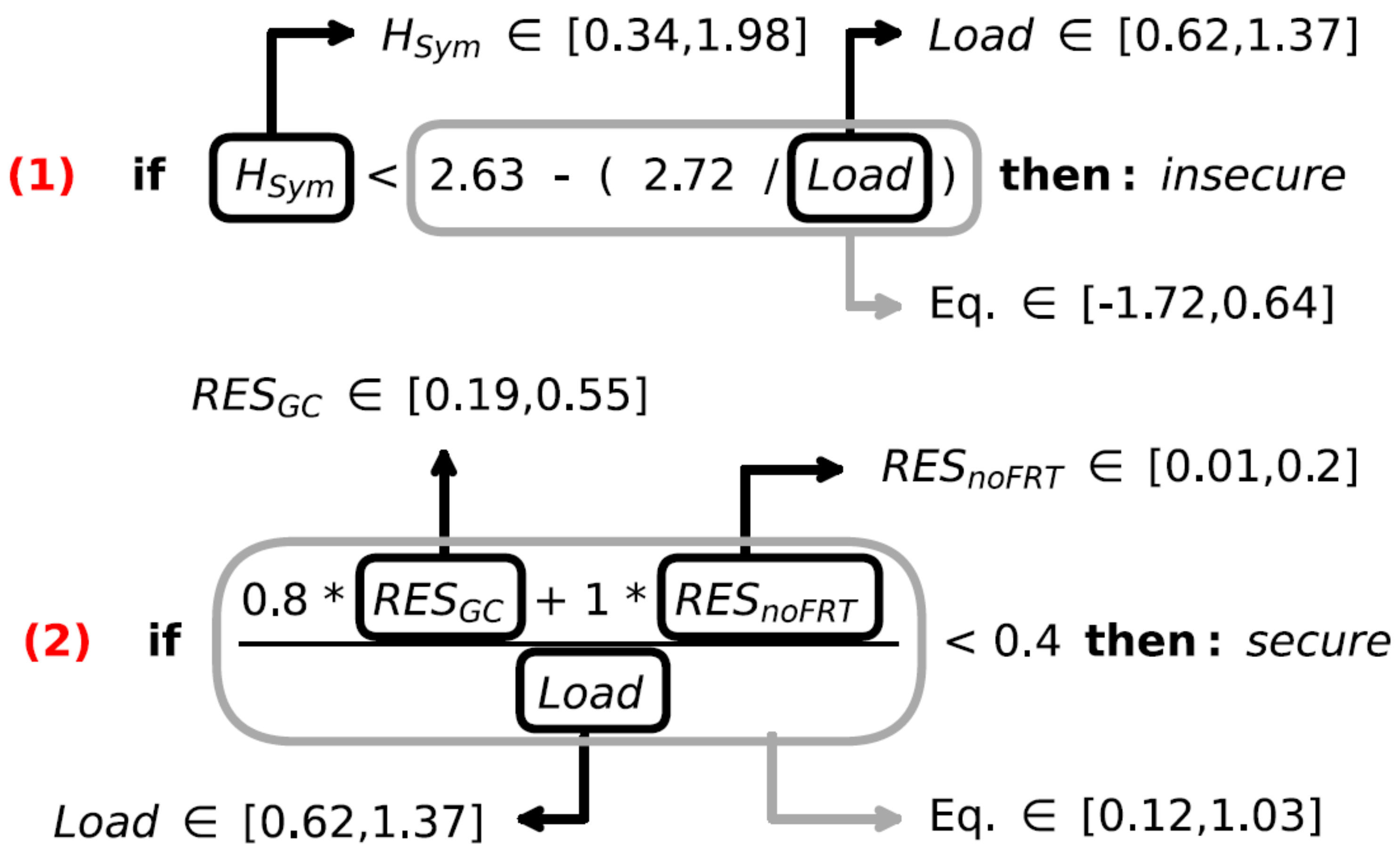
elif $\{ -7.87 \times Load + 3 \times Load \times H_{Sym} < -8.18 \}$:

return {1}

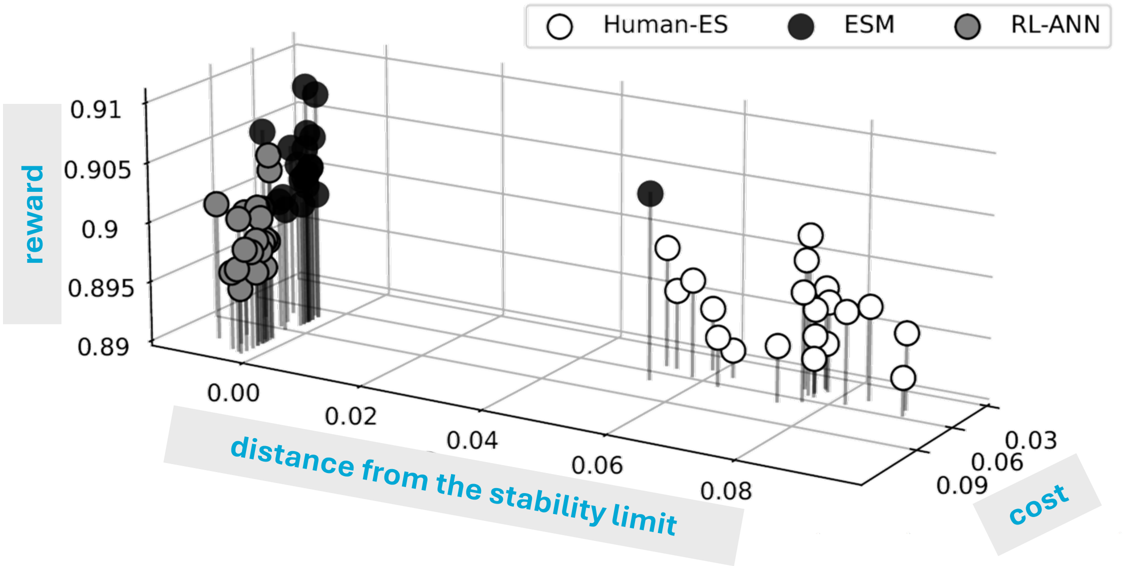
else :

return {0}

Physical meaning of the ESM

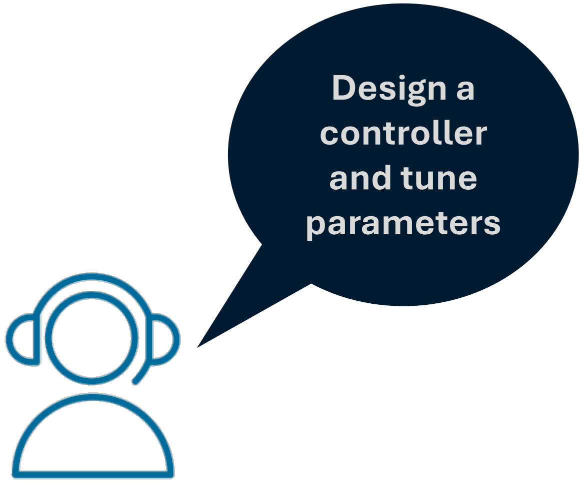


Results: Decision quality

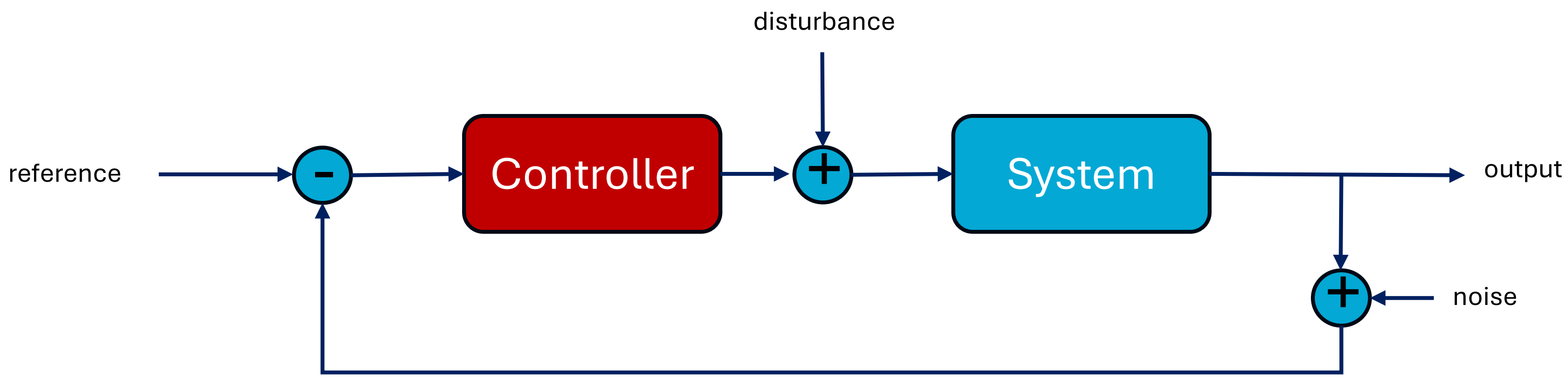


Human-ES: Human expert system
RL-ANN: Reinforcement learning artificial neural network

(2) Application to design controllers in power systems



Graph

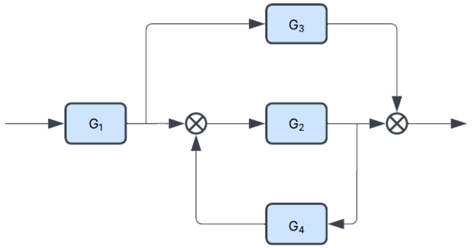


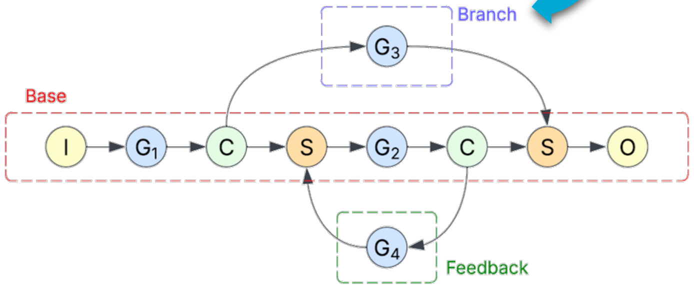
Reference: L. Bost, F. Fernandes, R.J. Bessa, "AI-assistant for intelligent design of controllers in power systems," *Sustainable Energy, Grids and Networks*, vol 43, pp. 101831, Sept. 2025.

ESM to design controllers

graph evolution process
using **Simulated Annealing**

Knowledge representation
(control graphs – *based on signal-flow graph*)

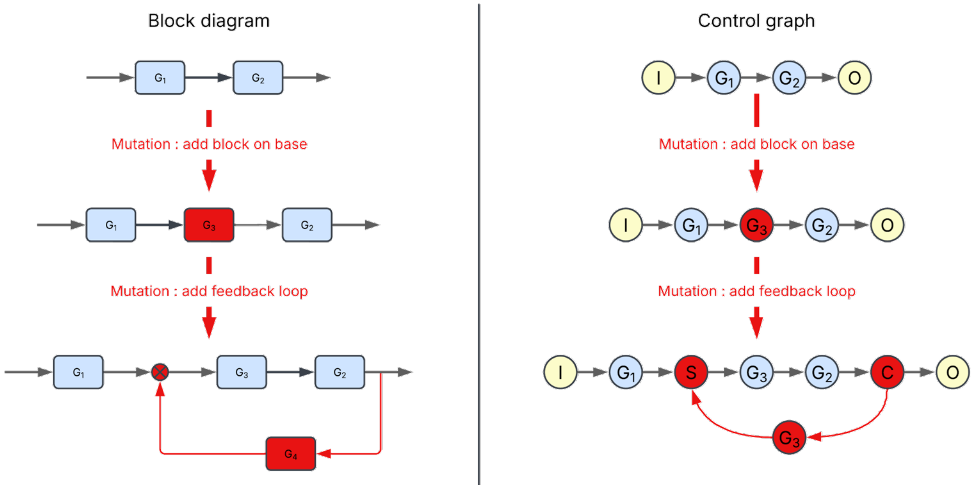




Algorithm 1 Control graph evolution process

Require: K_b, T_0, k_{max}

$k \leftarrow 1$
$T \leftarrow T_0$
Generate the first version of the control graph, CG
Compute $E = energy(CG)$
while $k < k_{max}$ **do**
  Generate neighbor control graph, CG_n (2.3)
  Tune numerical constants of CG_n (PSO)
  Compute $E_{new} = energy(CG_n)$
  if $p(E, E_{new}, T, K_b) \geq \mathcal{R}(0, 1)$ **then**
    $CG = CG_n$
  end if
  Adjust T according to cooling scheme
  Perform periodical simplification on CG
  $k \leftarrow k + 1$
end while
Perform simplification on CG



List of mutations

Code	Mutation	Parameters
1	Change transfer function on base	id, new function
2	Change transfer function out of base	id, new function
3	Add block on base	id, new function
4	Remove block on base	id
5	Add branch block	link, new function
6	Remove branch block	id
7	Add feedback block	link, new function
8	Remove feedback block	id

ESM to design controllers

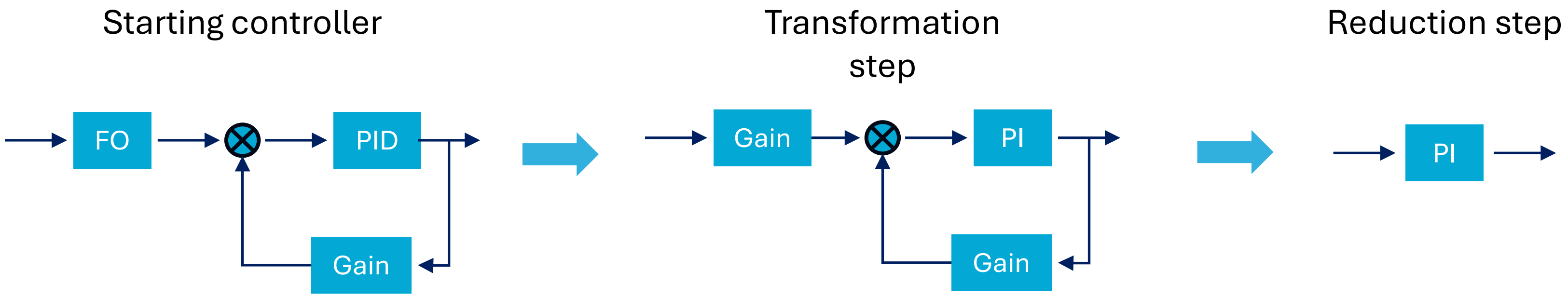
Step 1: List of individual block simplification

Original block	Simplified block	Condition
PI	Gain	$K_i = 0$
PD	Gain	$K_d = 0$
PID	PI	$K_d = 0$
PID	PD	$K_i = 0$
PID	Gain	$K_i = 0,\ K_d = 0$
First order	Gain	$\tau = 0$
Second order	Gain	$\omega_n = 0$
Lead	Gain	$\tau = 0$
Lag	First order	$\alpha = 0$
Lag	Gain	$\tau = 0$
SM	Gain	$H = 0$

Algorithm 1 Control graph evolution process

Require: K_b, T_0, k_{max}

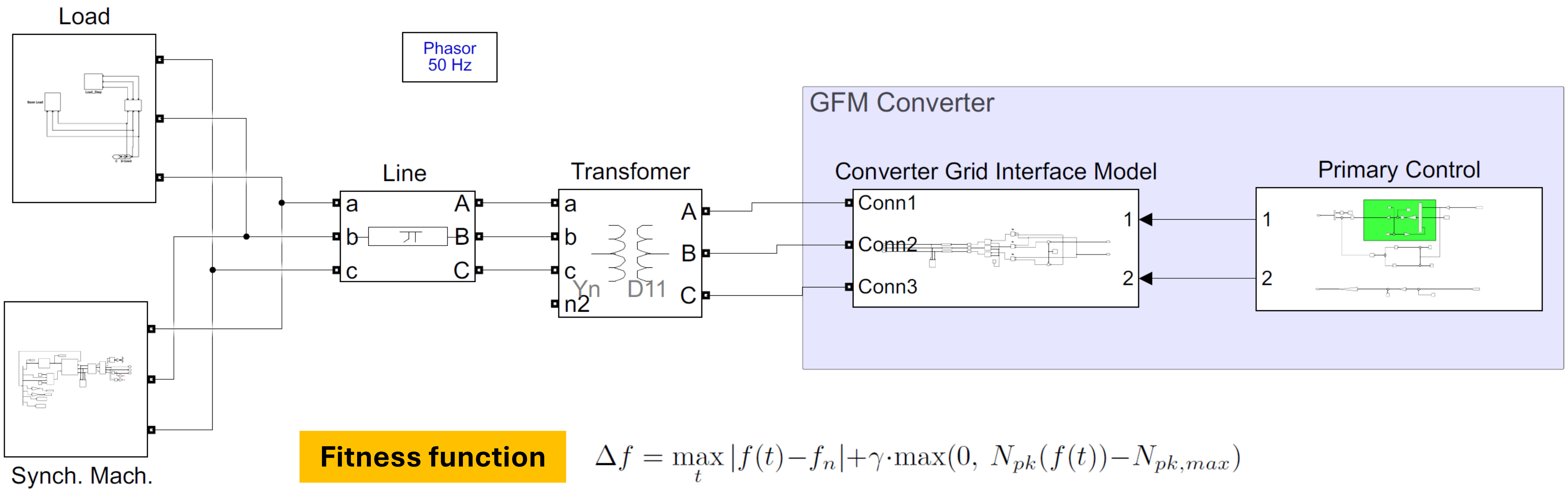
```
k ← 1
T ← T_0
Generate the first version of the control graph, CG
Compute E = energy(CG)
while k ≤ k_max do
    Generate neighbor control graph, CG_n (2.3)
    Tune numerical constants of CG_n (PSO)
    Compute E_new = energy(CG_n)
    if p(E, E_new, T, K_b) ≥ R(0, 1) then
        CG = CG_n
    end if
    Adjust T according to cooling scheme
    Perform periodical simplification on CG
    k ← k + 1
end while
Perform simplification on CG
```



Step 2: Involves removing gain blocks with a value of K = 0 outside the base (branch and feedback blocks) and gain blocks with a value of K = 1 on the base

Grid-forming converter case study

Goal: generate a controller that improves the frequency stability

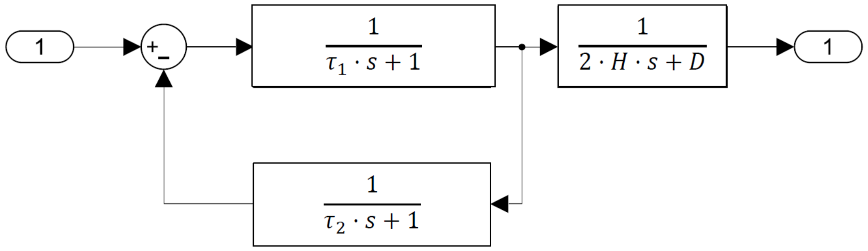


Fitness function

$$\Delta f = \max_{t} |f(t) - f_n| + \gamma \cdot \max(0,\ N_{pk}(f(t)) - N_{pk,max})$$

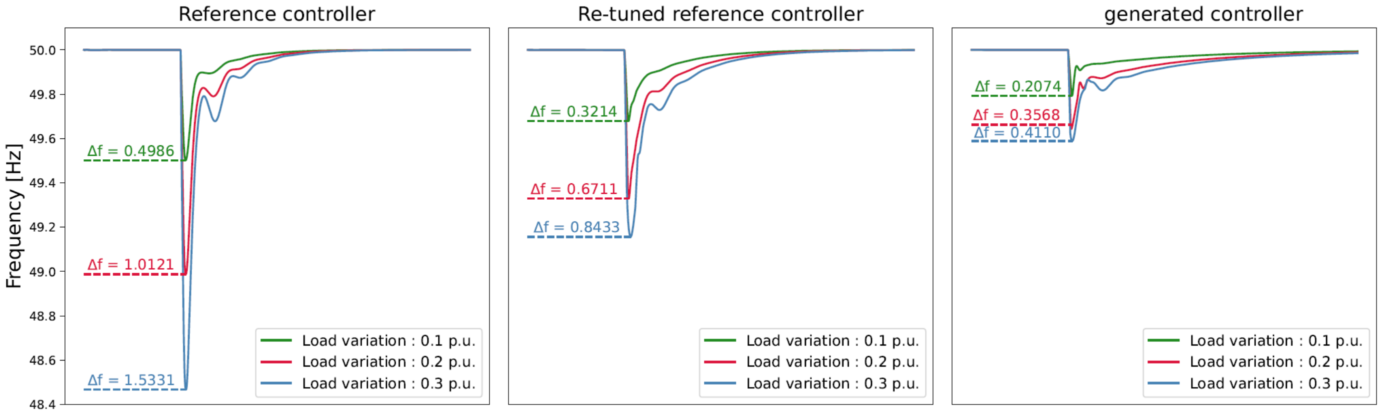
- Frequency deviation relative to its nominal value + penalty that discourages highly oscillatory responses
- During parameter tuning (PSO), the stability margins criterion is used as the cost function of the PSO, and constructed by linearizing the model, and the open-loop transfer function of the system is extracted

Results

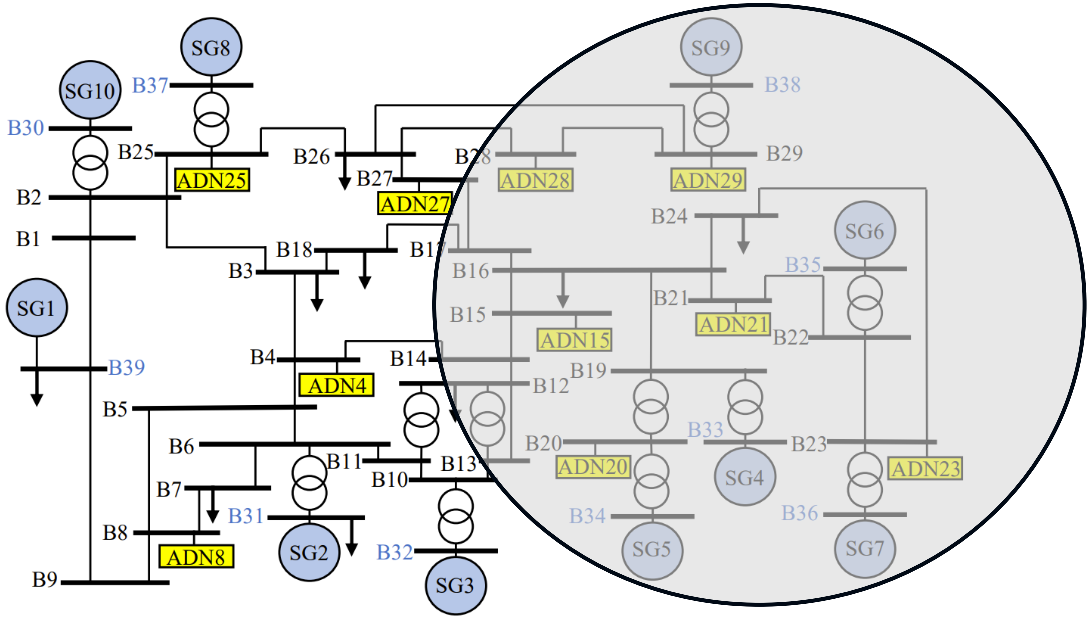


Control system learned with ESM for the GFM converter

benchmark models



(3) Explaining power systems dynamics



What variables (and how) affect the system stability at a local level?

Important to

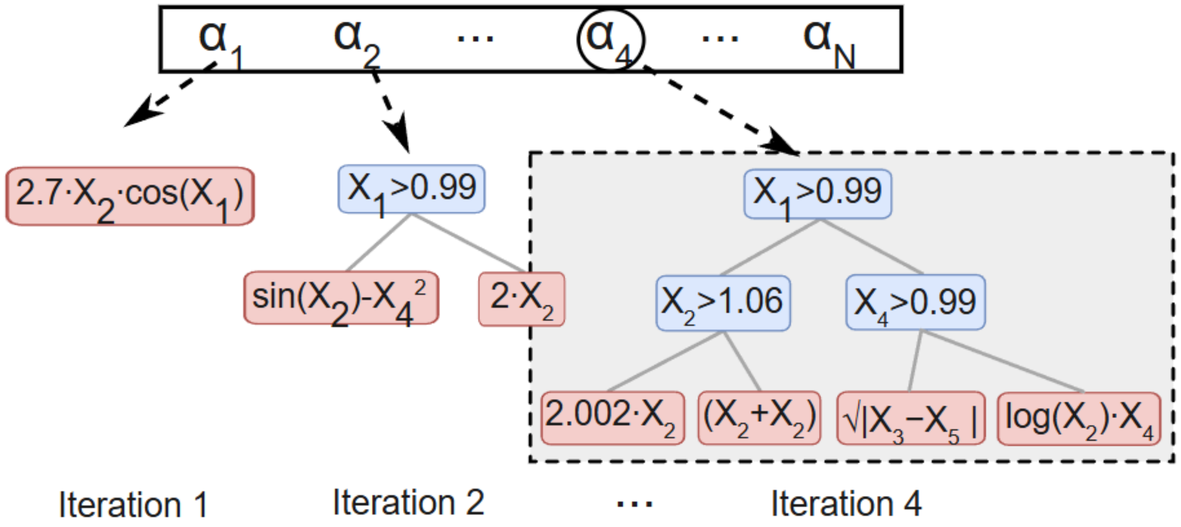
- Enhance situational awareness for day-to-day operation
- Help in defining preventive measures to avoid stability or security loss

Derive the critical clearing time (CCT) expression considering a 3-phase self-clearing fault at an HV bus

Reference: I. Lamprianidou, F. Fernandes, R.J. Bessa, P. Papadopoulos, "Symbolic explainer of power system dynamics," *PSCC 2026, Limassol, Cyprus, June 2026*.

Piecewise Symbolic Regression (Pc-SR)

(1) Iterate along CCP path (characterized by $\{a_k\}$):



(2) Post-Hoc Merging

1) Canonicalization of expressions:

$(X_2+X_2) \rightarrow 2 \cdot X_2$

2) Merging based on Embedding Similarity:

$(X2 \cdot 2.002 \; ; \; 2 \cdot X_2) \rightarrow 2 \cdot X_2$

(3) Interpretable rule based model

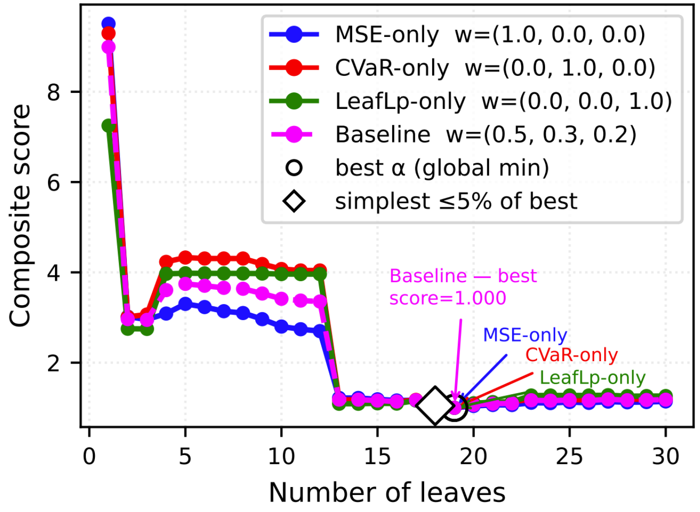
if ($X_1 > 0.99$) **:** $2 \cdot X2$
elif ($X_4 > 0.99$) **:** $\log(X2) \cdot X4$
else : $\sqrt{|X3 - X5|}$

Results

Deriving RC CTT equations for Bus 29

TABLE III
GLOBAL PERFORMANCE METRICS OF PC-SR.

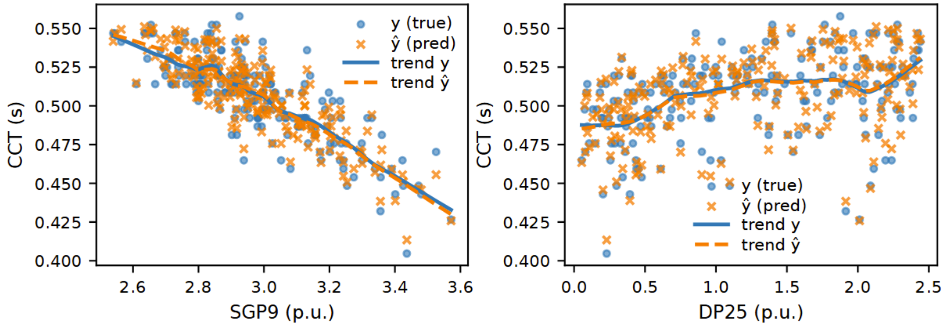
Metric	R^2	MSE (s^2)	RMSE (s)	$CVaR_{95}$ (s^2)
Score	0.999	2.3e-05	0.0047	1.6e-04



Example of resulting critical clearing time expression and corresponding region

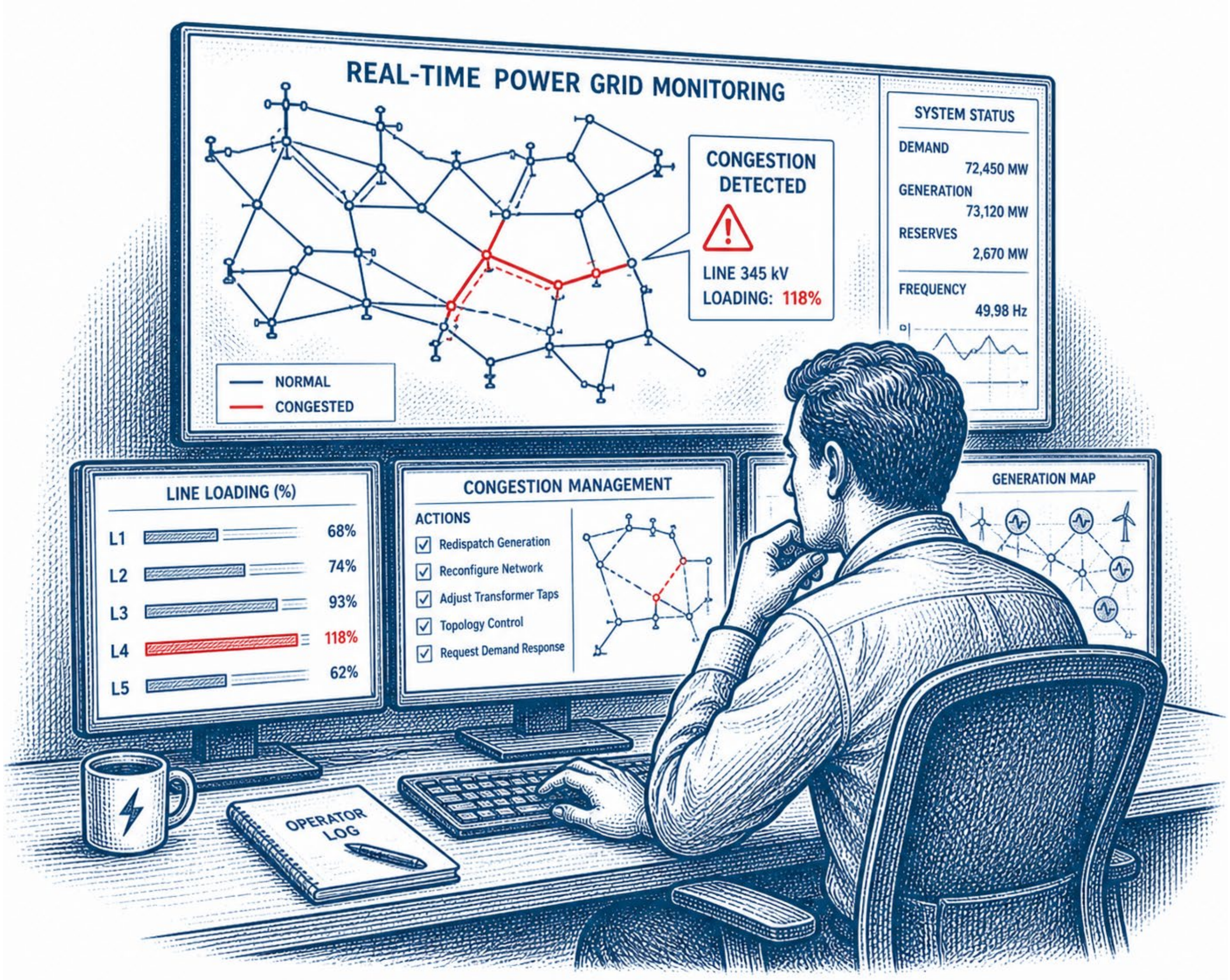
$$\mathcal{R}_{18}^{\alpha^\dagger} = \left\{ x \in \mathbb{R}^d : \begin{array}{l} SGKE5 > 8.12,\ SGKE8 > 21.26, \\ SGP6 > 3.79,\ SGP7 > 2.88 \end{array} \right\}$$

$$g_{18}^{\alpha^\dagger}(x) = 0.012167\,(\delta_5 \cdot DP25) + \frac{\delta_1 + 1.3266}{\delta_9 + SGP9}.$$



(a) CCT vs. $SGP9$ (b) CCT vs. $DP25$

(4) Real-time congestion management

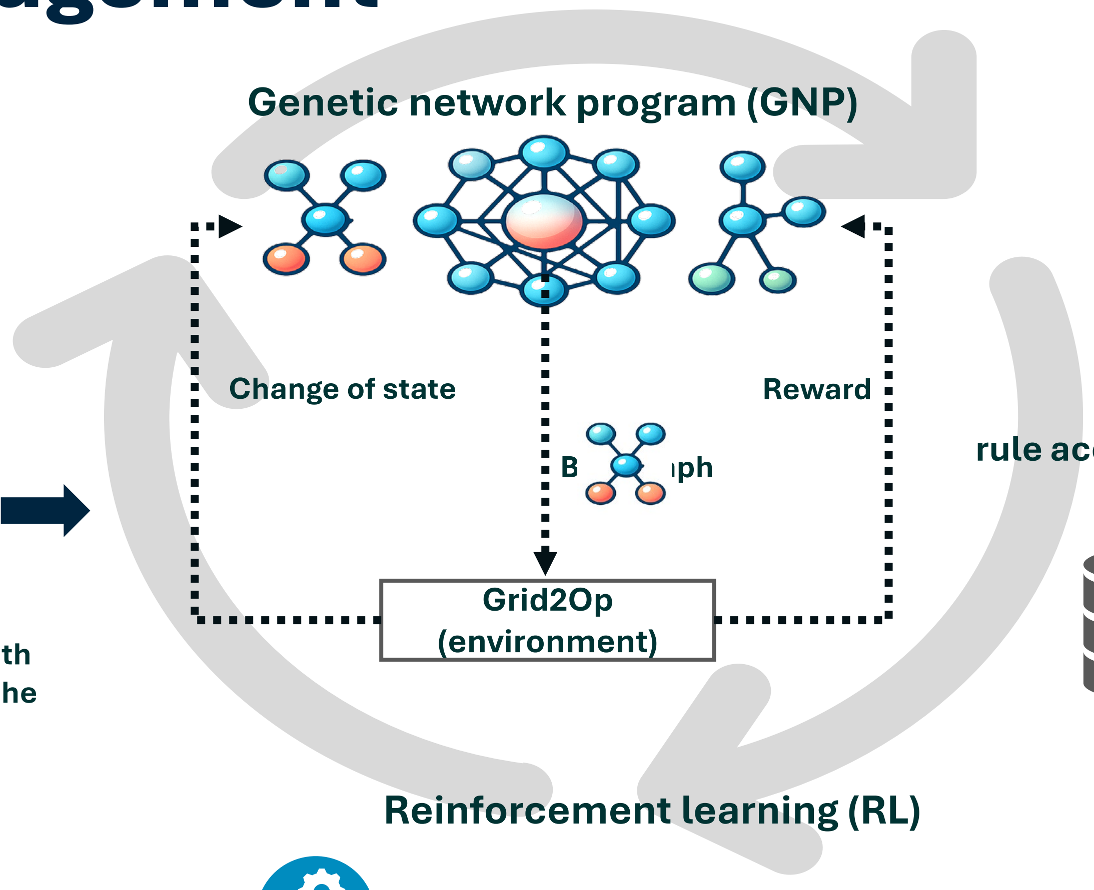


[generated with ChatGPT]

Evolving rules for real-time congestion management

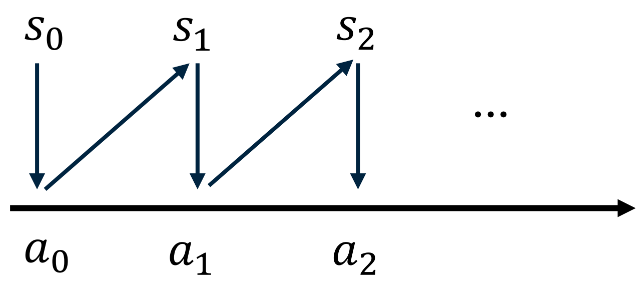
expert system

initializing AI agent with modified versions of the expert decision graph



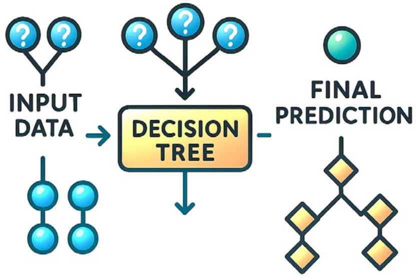
GNP will extract the best-evolved (elite) decision graph per episode

rule accumulation



Accumulating optimized rules via following the trajectory of actions $(a_0, a_1, .., a_n)$ selected per grid sate $(s_0, s_1, .., s_n)$ by elite graphs

rule extraction by a decision tree

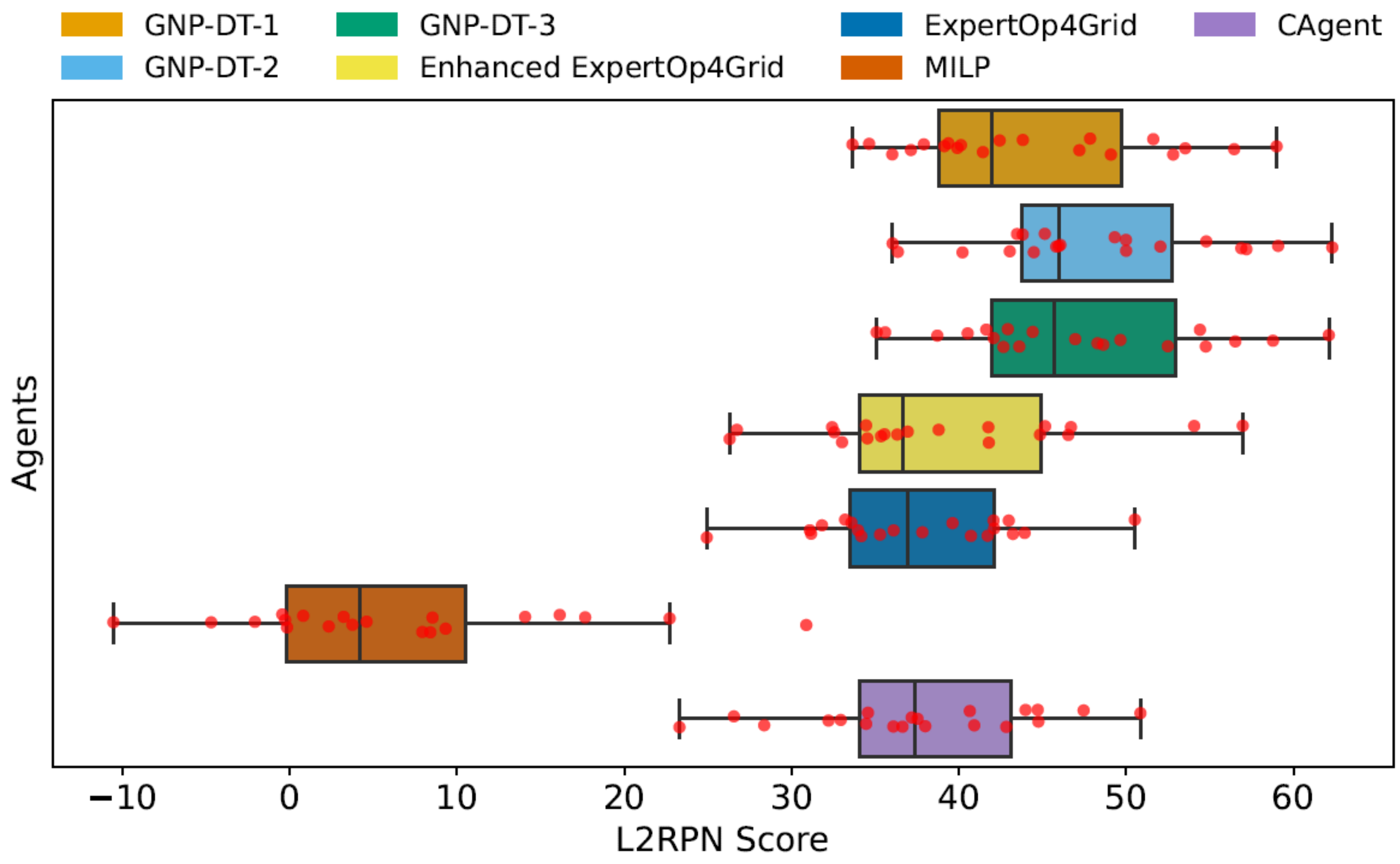


AI4 REALNET

INESC TEC

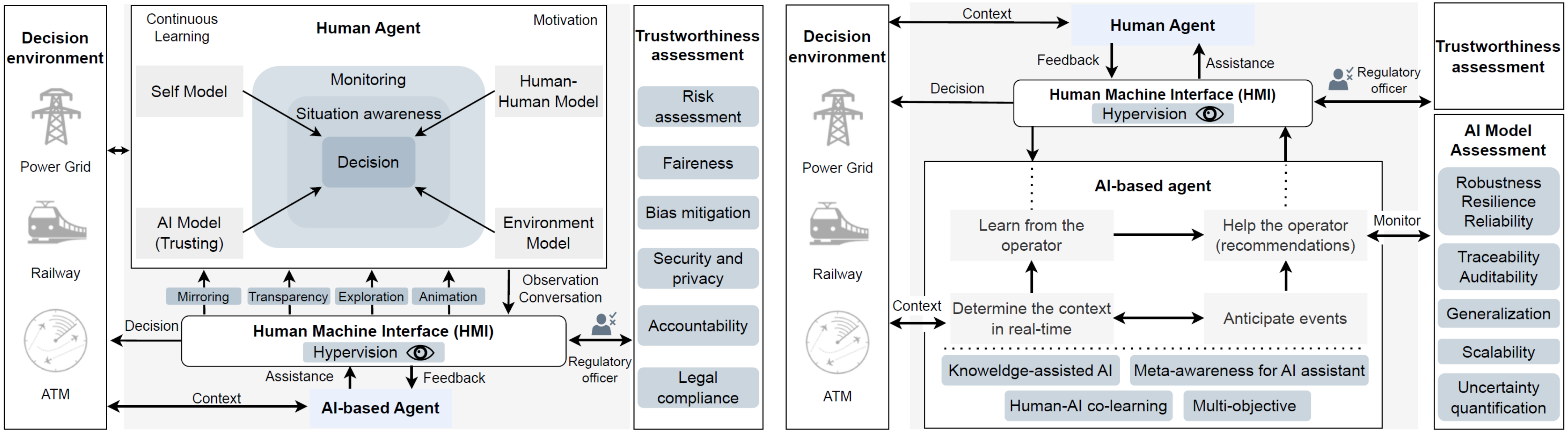
Reference: F. Moaidi, R.J. Bessa, “Evolving power system operator rules for real-time congestion management,” *Energy and AI*, vol. 23, pp. 100672, Jan. 2026.

Results for IEEE 118 (Grid2Op environment)



AI assistants for power grids

INESC TEC



(a) Human agent decision-making

(b) AI supported decision-making

Reference: M. Mussi, A. Maria Metelli, M. Restelli, G. Losapio, R. J. Bessa, et al., "Human-AI interaction in safety-critical network infrastructures," *iScience*, vol. 28, no. 9, pp. 113400, Sept. 2025.

Reference: M. Leyli-abadi, R.J. Bessa, et al., "A conceptual framework for AI-based decision systems in critical infrastructures," *2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Vienna, Austria, 5-8 October 2025.

Human-AI co-working interfaces

Bi-directional virtual assistants for joint decision-making | Hypervision concept

notifications about risks and events

Alerts

Overload on line 5_10_7

Warning: the line 7 connecting nodes 5 and 10 is overloaded by 112%

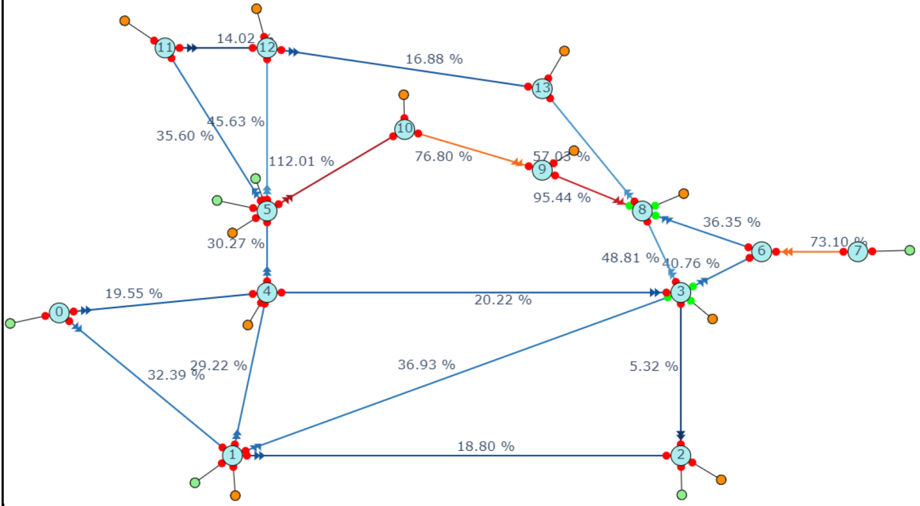
Security

Risk of N-1 contingency on line 9_10_12

Security

real-time view of the environment with tools like zooming

Context



AI-based suggestions that operators can adopt

Recommendations

Topological change

Switching to the scheme at substation 5

Redispatch

Reduce the generator 2 by 3 MW

Timeline

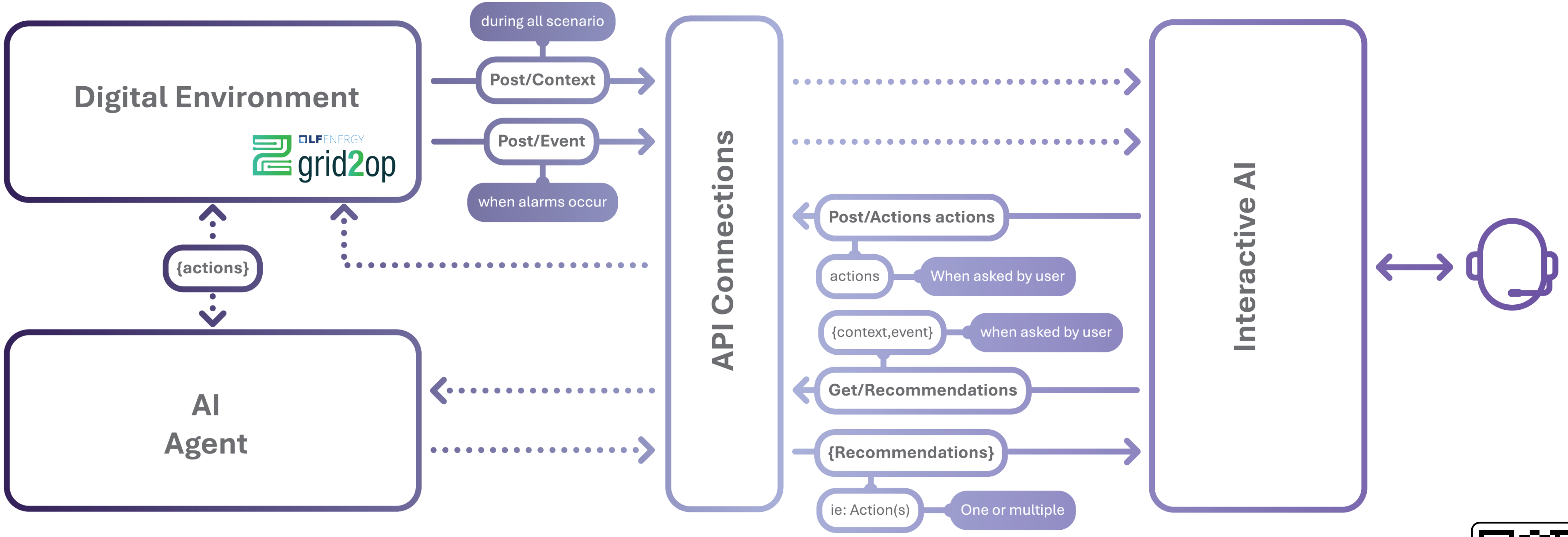
16:00 16:10 17:00 18:00 19:00

Overload on line 5_10_7

16:10 16:15

tracking time steps and event history for analysis

AINETUS – Linux Foundation Energy project



CALL FOR CONTRIBUTORS

AINETUS demo

National Aeronautics and Space Administration

System Failure Case Studies

December 2007 Volume 1 Issue 10

Powerless

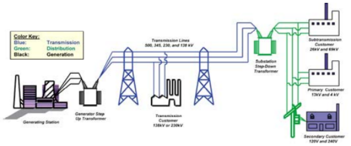
On August 14, 2003, the United States and Canada experienced the largest electrical power blackout in North American history. It was a massive power outage that affected parts of the northeastern U.S. and eastern Canada. Approximately 40 million people in eight U.S. states (about one-seventh of the population of the U.S.) and 10 million people in the Canadian province of Ontario (about one-third of the population of Canada) were impacted. The cost of financial losses related to the outage was estimated at $4 to $10 billion. The shutdown was the result of a monitoring and diagnostic systems failure coupled with communications problems between operations and support staffs, and a lack of systems understanding and planning by utility operators.

Background: "The Grid"

The North American power grid is one large, interconnected system, considered to be one of the greatest engineering achievements of the past 100 years. Its infrastructure is valued at more than $1 trillion, with more than 200,000 miles of transmission lines operating at 230,000 volts and greater, 950,000 megawatts of generating capability, and 3,500 utility organizations serving well over 283 million people.

The electrical power system or grid produces electricity from fuel sources, such as nuclear, coal, oil, natural gas, hydro power, geothermal, etc. Low voltage electricity from the generators (10,000 - 25,000 volts) is "stepped up" to higher voltages (230,000 - 765,000 volts) for transmission over power lines. Transmission lines are interconnected at switching stations and substations to form a network. Electricity flows through the network following the laws of physics—along "paths of least resistance," the same way that water flows through a network of canals. When the power arrives near a load center, it is stepped down to lower voltages for distribution to residential customers (120 and 240 volts) or larger industrial and commercial customers (12,000 - 115,000 volts).

Electrical power cannot easily be stored over extended periods of time, and is consumed immediately after being generated.



Basic Structure of the Electric System.

The demand load on any power grid must be matched by its supply and ability to transmit that power. Any significant overload of a power line or underload/overload of a generator requires utilities to disconnect the line or generator from the grid to prevent hard-to-repair and costly damage.

Although the power system in North America is commonly referred to as the grid, it is actually a group of three distinct power grids or that are electrically independent from each other. They are: the Eastern Interconnection, which includes the eastern two-thirds of the continental U.S. and Canada; the Western Interconnection; and the state of Texas.

In August of 2003, the largest blackout in North America occurred, affecting 50 million people at an estimated cost of $4 - $10 billion

Proximate Causes:

- Load imbalance caused by generator shutdown triggered cascading transmission line failure

Underlying Issues:

- Poor communication of software failures
- Inadequate system planning and understanding
- Tree overgrowth near high voltage lines
- Lack of thorough operator training

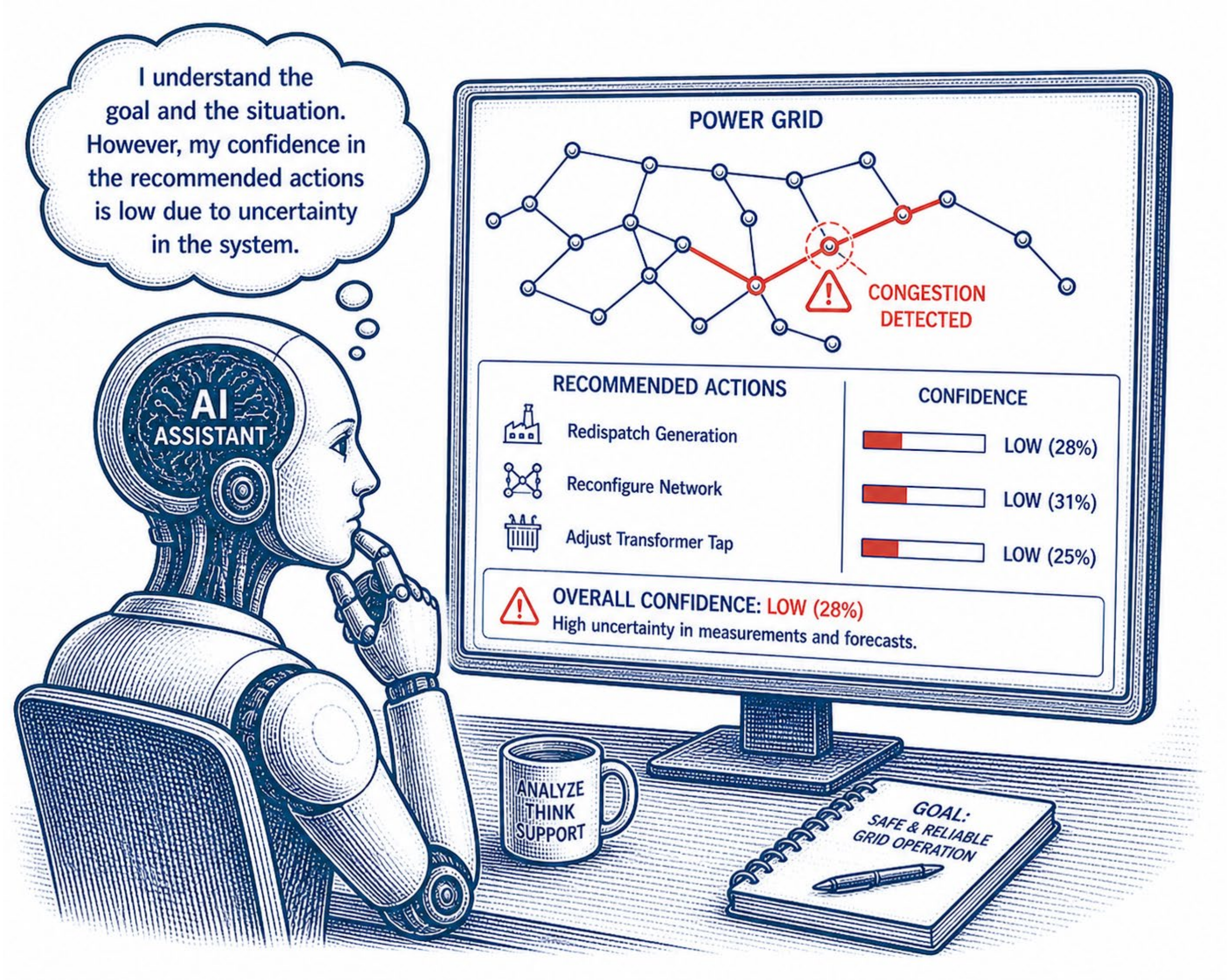
*"…**software timing error** in FirstEnergy's UNIX-based XA/21 **energy management computer was found to be the primary cause of the grid event alarm failure**. After the alarm system failed silently, the unprocessed events started to queue up and crashed the primary server within 30 minutes.*

*"**The operators only determined they had problems when data from phone calls received from customers**, nearby utilities, and their regional coordinating center calls did not match the information on their screens."*

"The blackout might have been prevented if FirstEnergy's operators only knew what was happening with their grid"

What if the failure was in an AI assistant?

Self-awareness of the AI assistants

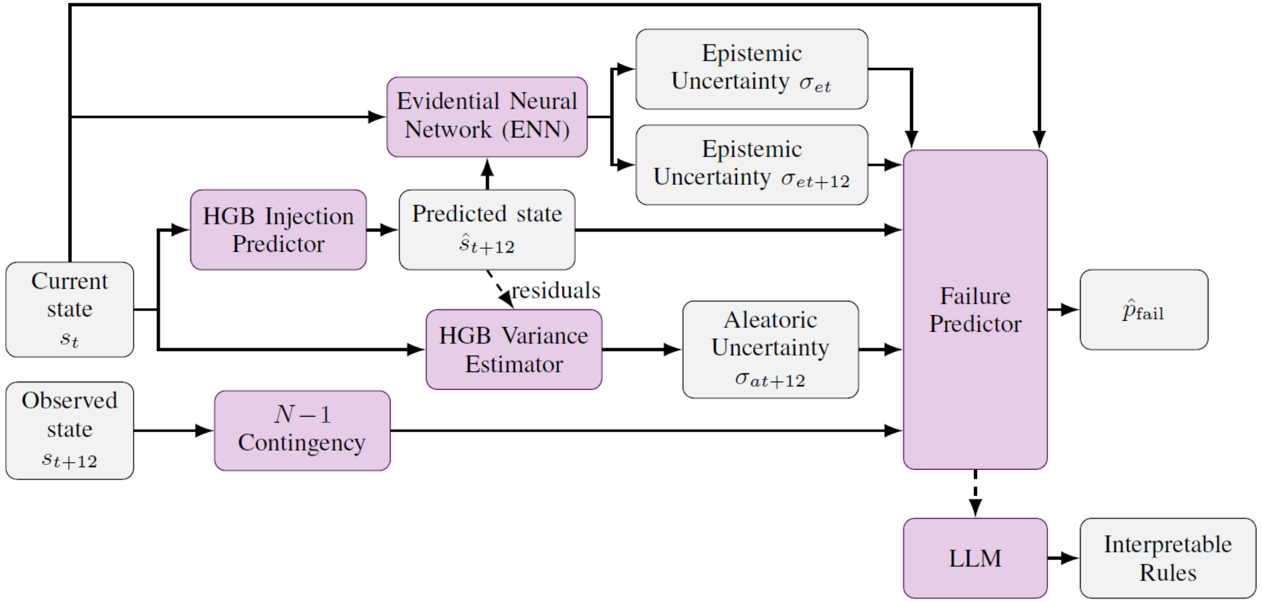


[generated with ChatGPT]

Machine learning method that **integrates epistemic and aleatoric uncertainty**, and indicators of power grid operational complexity, to **predict the probability of reinforcement learning agent failure** over future time horizons

⇨ **used to defer the "control" to humans**

AI agent failure prediction



Performance for one-hour ahead reinforcement learning agent failure prediction

Configuration	F_2-score	Accuracy	OVR	FA
Baseline (grid variables)	70.81%	85.67%	7.65%	14.95%
Epistemic-only	35.40%	42.71%	20.06%	60.78%
Grid + aleatoric	73.62%	87.57%	7.08%	12.93%
Grid + aleatoric + epistemic	**76.53%**	**89.51%**	**6.79%**	**10.84%**

FA – False Alarms
OVR – Oversight Rate

Reference: R. Bessa, M. Costa, C. Gonçalves, P. Ferreira, "Uncertainty quantification and failure Prediction for AI-based congestion management," *submitted to SEST 2026*.

INESC TEC

Featured 3 Feb 2025

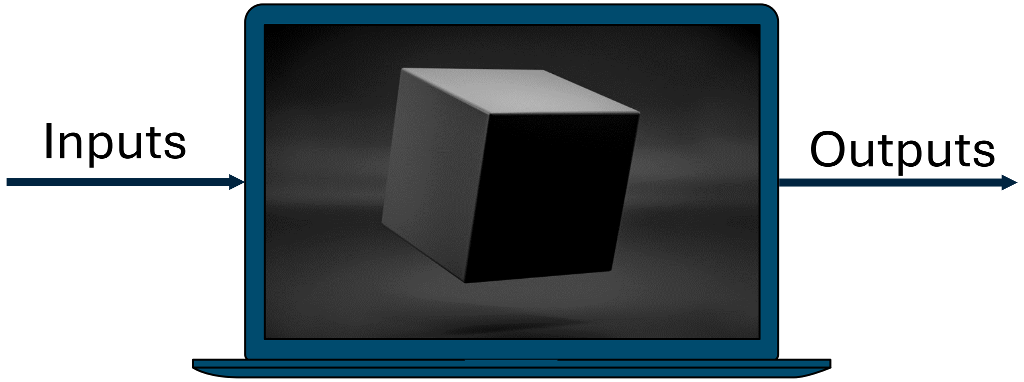
Flow-based market coupling & counterintuitive flows in Europe

8 min

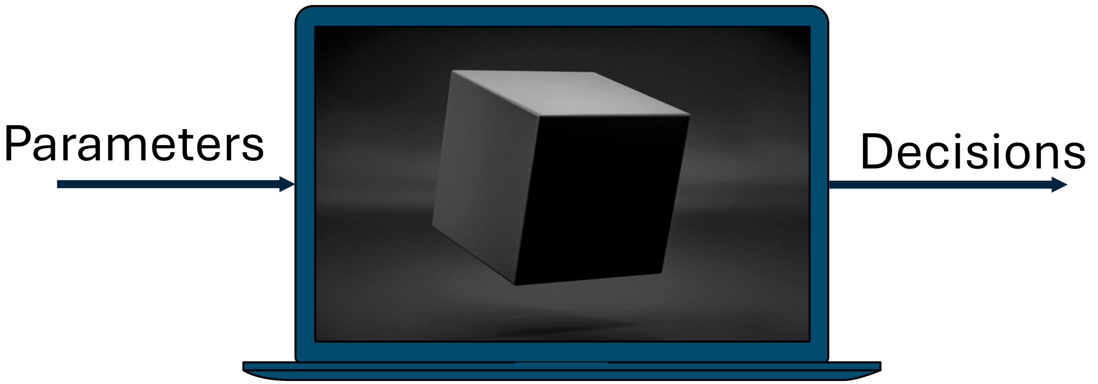
We look into the key drivers behind counterintuitive flows and examines the operational challenges and opportunities they create for market participants.

Su

machine learning is frequently seen as a “black box”



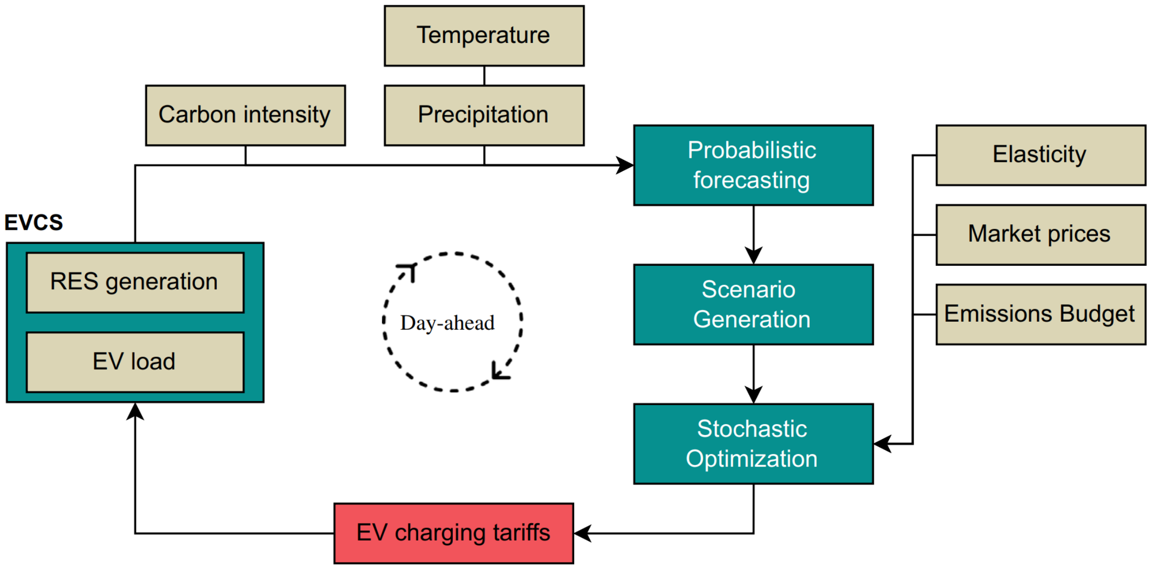
but, **mathematical optimization** can also be **seen as a “black box” by a human** decision-maker

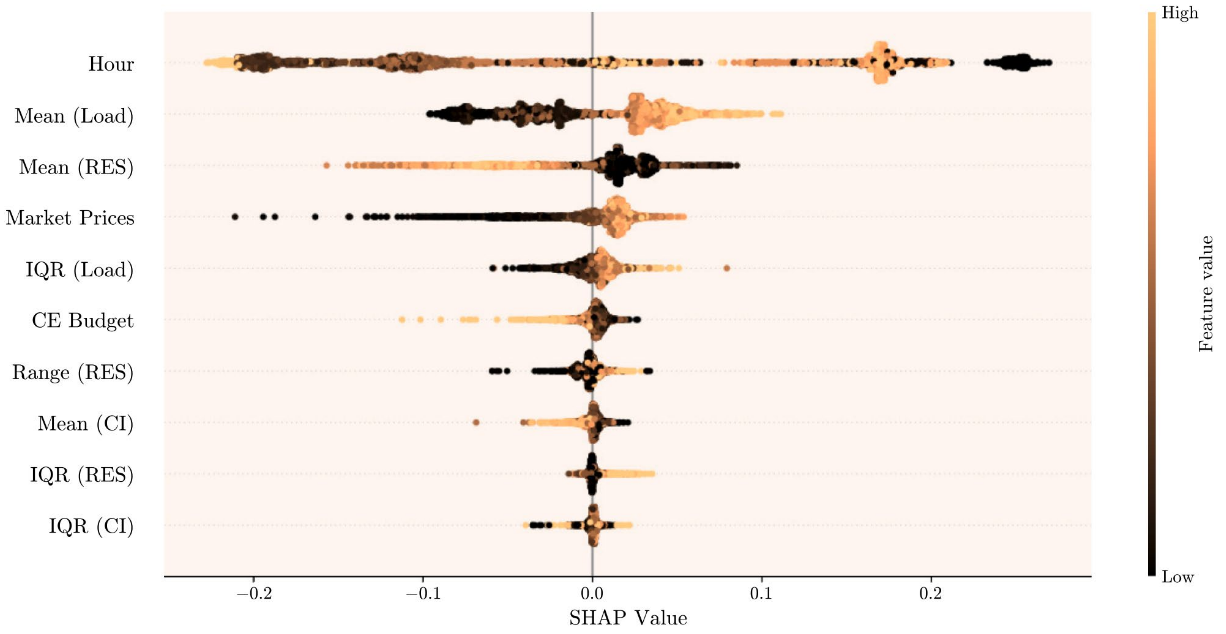


especially if there is forecast uncertainty

Explaining EV charging tariffs design

Explain the results of an optimization problem to a decision-maker ▷ **design of EV charging dynamic tariffs**

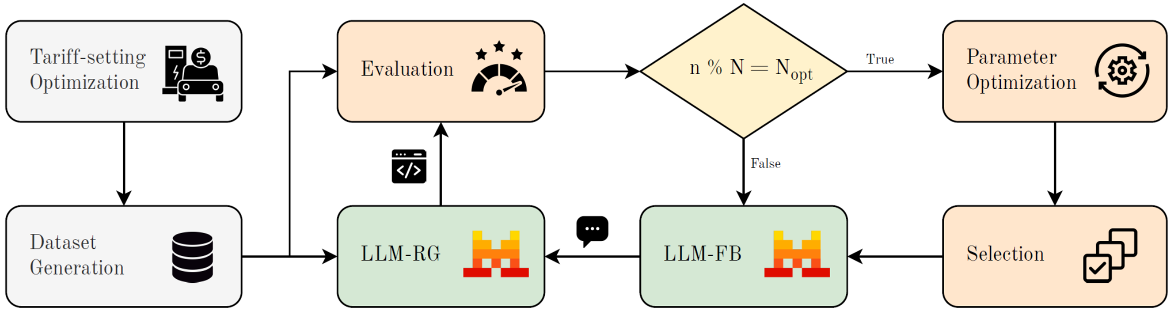




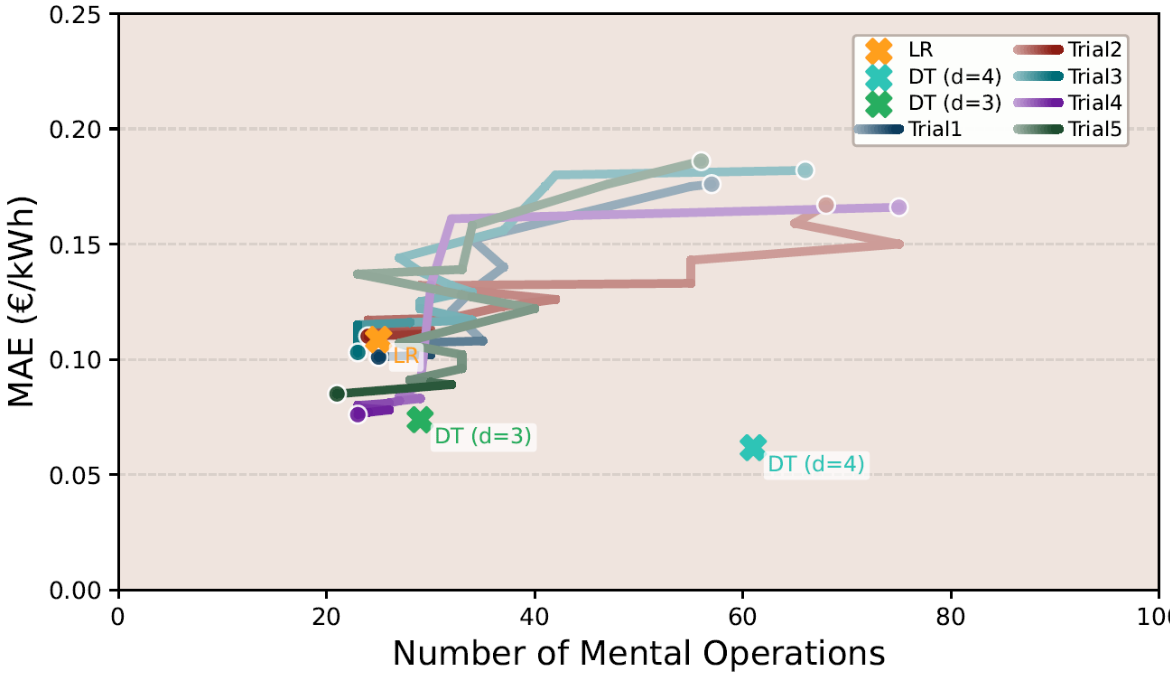
Is feature importance really meaningful?

Reference: C. Silva, R.J. Bessa, "Carbon-aware dynamic tariff design for electric vehicle charging stations with explainable stochastic optimization," Applied Energy, vol. 389, pp. 125674, July 2025.

Explaining EV charging tariffs design



imitation accuracy



Global interpretability: total number of mental operations required to inspect the complete model

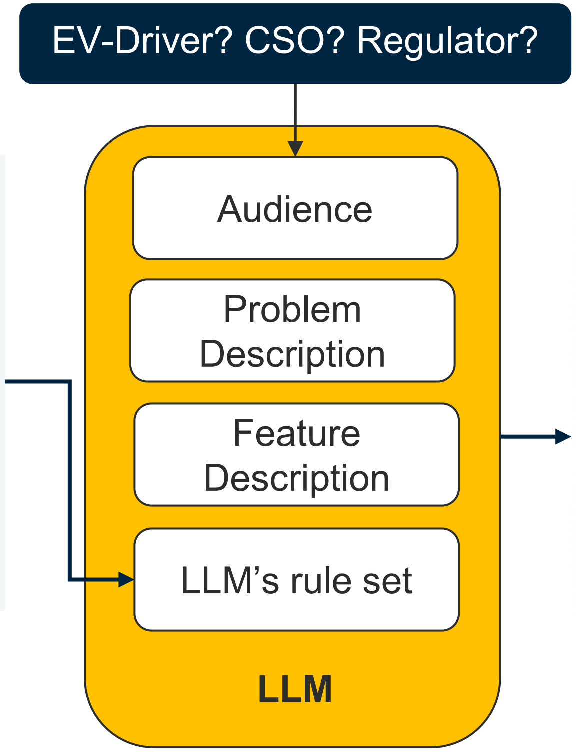
Reference: T. Bessa, C. Silva, R.J. Bessa, "Explaining the optimization of dynamic tariffs for electric vehicles charging with large language models," *submitted to SEST 2026*.

Explaining EV charging tariffs design

Symbolic Rules

```
def predict(features, X):
    mean_res = features['Mean (RES)']
    hour = features['Hour']

    if hour < 8.932:
        return max(0.05, 1.254 - (hour * 0.128) - 0.571)
    elif mean_res + (hour * -0.078) > 124.785:
        return max(0.261, min(1.243, -0.324 + (mean_res *
        ↪  -0.182)))
    elif hour >= 14.95:
        return max(0.105, min(0.6, 2.278 + (hour - 15.039)
        ↪  * -0.277))
    else:
        return max(-0.338, min(0.374, 0.234 + (mean_res *
        ↪  0.077 if mean_res < 49.999 else 0.117)))
```



Human-readable Explanation

```
{
"audience": "EV-User",
"summary": "Big takeaway: this rule mainly cares about the time of
day, with renewable supply only nudging the price around the middle
of the day. In practice, charging is usually cheapest very early in
the morning (it quickly drops to about 0.05EUR/kWh by around 5 AM),
usually mid-priced late morning to mid-afternoon (mostly around
0.35-0.37EUR/kWh), and usually most expensive from about 3 PM to 9 PM
(typically 0.60EUR/kWh). There is one exception: if renewable supply
is unusually high around late morning or early afternoon, the price
drops to a flat 0.261EUR/kWh. Very late at night, after the evening
peak, it falls again and can get down to about 0.105EUR/kWh near 11
PM. So, for a practical rule of thumb: cheapest is early morning,
worst is late afternoon/evening, and midday is usually in the middle
unless renewables are especially strong."
}
```

Reference: T. Bessa, C. Silva, R.J. Bessa, "Explaining the optimization of dynamic tariffs for electric vehicles charging with large language models," *submitted to SEST 2026*.

Interpretable AI agent failure prediction

INESC TEC

Algorithm 1 Dual LLM-based RL failure prediction

Require: Dataset $\mathcal{D}$, HGB model $\mathcal{HGB}$, Iterations (N)
Ensure: Best interpretable rule R_{best}

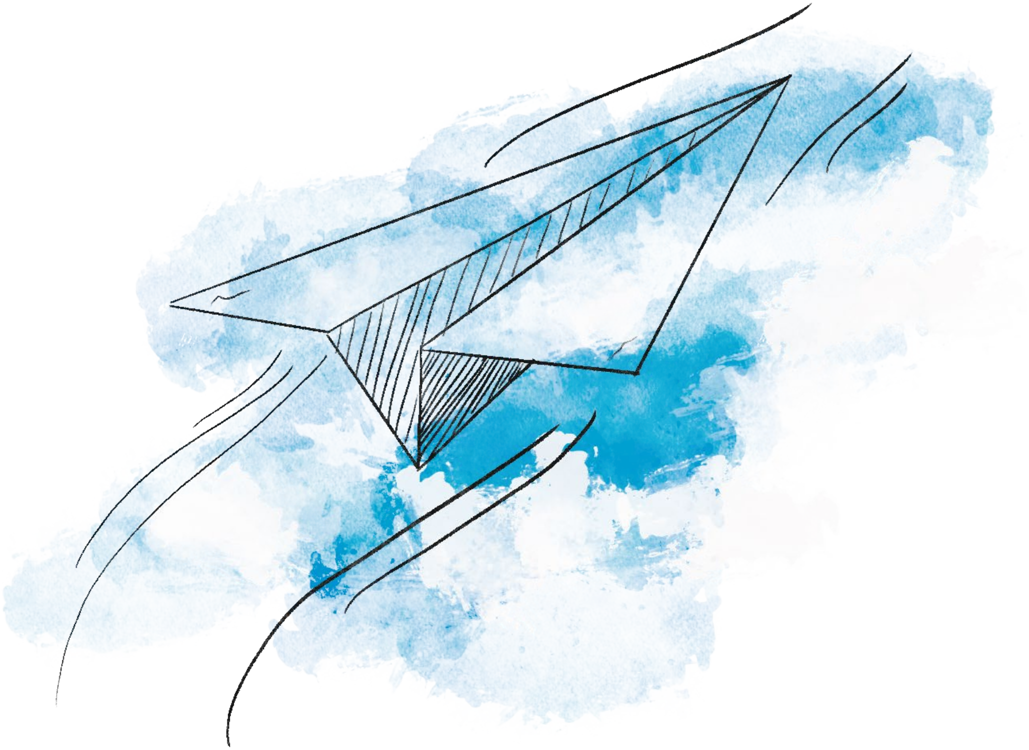
$Train, Test \leftarrow \text{SequentialSplit}(\mathcal{HGB}.Predict(\mathcal{D}))$
$Data_{bal}, Stats \leftarrow \text{SampleBalancedData}(Train)$
$R_{best}, S_{best} \leftarrow \text{Evaluate}(\text{InitHeuristicRules}(Stats), Test)$
$History \leftarrow \emptyset$, $StagCount \leftarrow 0$
for $iter = 1$ **to** N **do**
  $Osc \leftarrow \text{DetectOscillation}(History, \text{window} = 4)$
  $WC \leftarrow MisclassifiedCases(R_{best}, Test)$
  $R_{cand} \leftarrow \text{Gen}_{\text{LLM}}(Data_{bal}, Stats, WC, History, Osc)$
  if not $\text{Validate}(R_{cand})$ **then**
    $R_{cand} \leftarrow \text{Repair}_{\text{LLM}}(R_{cand})$
  end if
  $Metrics_i \leftarrow \text{ComputeMetrics}(R_{cand}(Test.X), Test.Y)$
  $S_i \leftarrow \text{ScoringFunction}(Metrics_i)$
  $Feedback \leftarrow \text{Critic}_{\text{LLM}}(R_{cand}, Metrics_i, Stats, Osc)$
  $History.\text{Append}(R_{cand}, Metrics_i, S_i, Feedback)$
  if $S_i > S_{best}$ **then**
    $R_{best}, S_{best} \leftarrow R_{cand}, S_i$
    $StagCount \leftarrow 0$
  else
    $StagCount \leftarrow StagCount + 1$
  end if
end for
return R_{best}, S_{best}

```python
def rule(x):
    if x["max_line_rho"] >= 0.65:
        if x["epistemic_before"] >= 0.7891:
            if x["aleatoric_gen_p_mean"] <= 0.3434:
                return 1
            else:
                return 0
        else:
            return 0
    else:
        if x["fcast_sum_load_q"] >= 155.5429:
            if x["aleatoric_gen_p_mean"] <= 0.3434:
                return 1
            else:
                return 0
        else:
            return 0
```

Following a contingency on line 34_35_110, the RL agent is predicted to fail if the maximum line loading (ρ) at t is ≥ 0.65 while the epistemic uncertainty is ≥ 0.79, and the mean aleatoric generation uncertainty is ≤ 0.34, or if the forecasted reactive load at t+12 is ≥155.54 MVar and the mean aleatoric generation uncertainty is ≤0.34

Reference: R. Bessa, M. Costa, C. Gonçalves, P. Ferreira, "Uncertainty quantification and failure Prediction for AI-based congestion management," *submitted to SEST 2026*.

Concluding remarks



- Interpretability (beyond explainability) is critical for real-world adoption
- Hybrid approaches (physics + data + optimization) will dominate over purely data-driven methods creating compound AI systems
- AI will enable faster, scalable, and more adaptive decision-making under uncertainty
- Human-AI collaboration will be key in control rooms and operational workflows
- Progress requires standardization, shared models/data, and interdisciplinary collaboration

WE ARE SCIENCE.
WE ARE TECHNOLOGY.
WE ARE INNOVATION.
WE ARE INESC TEC.

Work developed in

Smart4RES

ENFIELD

AI4 REALNET