Trustworthy AI Solutions for Cyberbiosecurity Challenges in Water Supply System

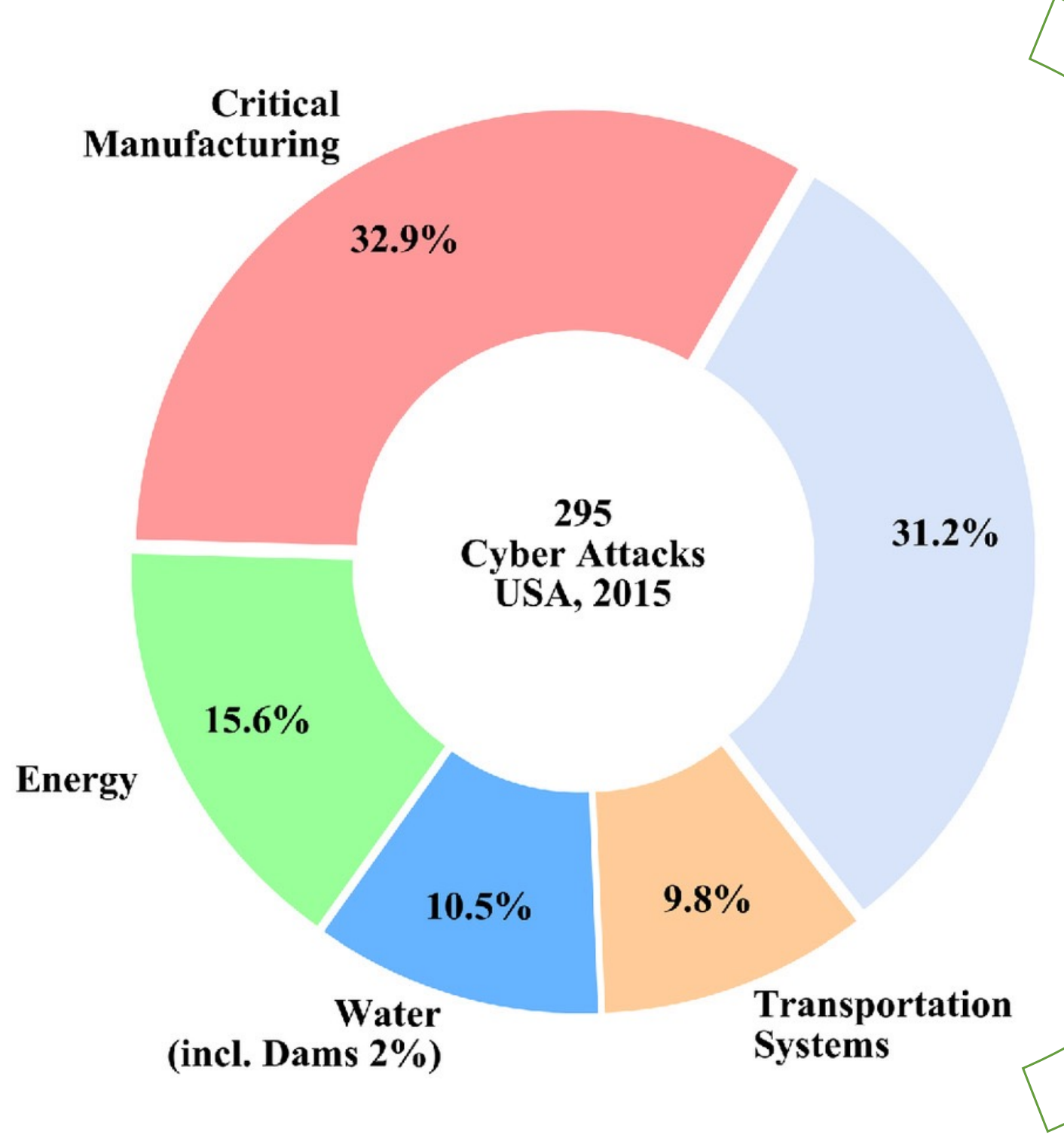
Wan-Yi Mao, Mehmet Yardimci, Minh T. Nguyễn, Dan Sobien, Laura Freeman, Abdul Rahman, Vinita Fordham, Feras A. Batarseh

{wanyi, oguzy, mnguyen0226, sdan8,laura.freeman}@vt.edu, {Abdulrahman, vfordham}@deloitte.com, batarseh@vt.edu

1. Abstract

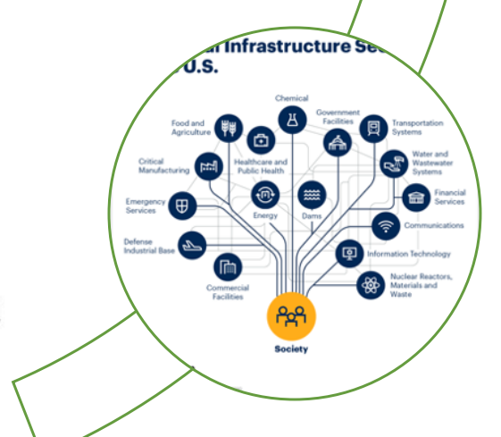
The recent increased adoption of AI into cyber-physical systems requires that AI models perform as intended and without security blunders. However, for such critical AI deployments, serious adversaries and data poisoning issues are still persistent. Accordingly, our research is dedicated to defining methods that mitigate different types of attacks and their undesired consequences. We are building an AI framework that can detect anomalies and malicious acts (in water supply systems) as well as analyze cause-effect measures in multiple scenarios. The proposed defense strategies include assessing the potential vulnerabilities that can be exploited in the convergence of AI and physical systems. Such outliers can cause extensive civilian harm, produce damaging bio-security incidents, and threaten agricultural and food production. The new framework will be applied to multiple physical systems with the aim of creating meta-learning outcomes for supporting the wide-scale adoption of trustworthy AI.

2. Motivation

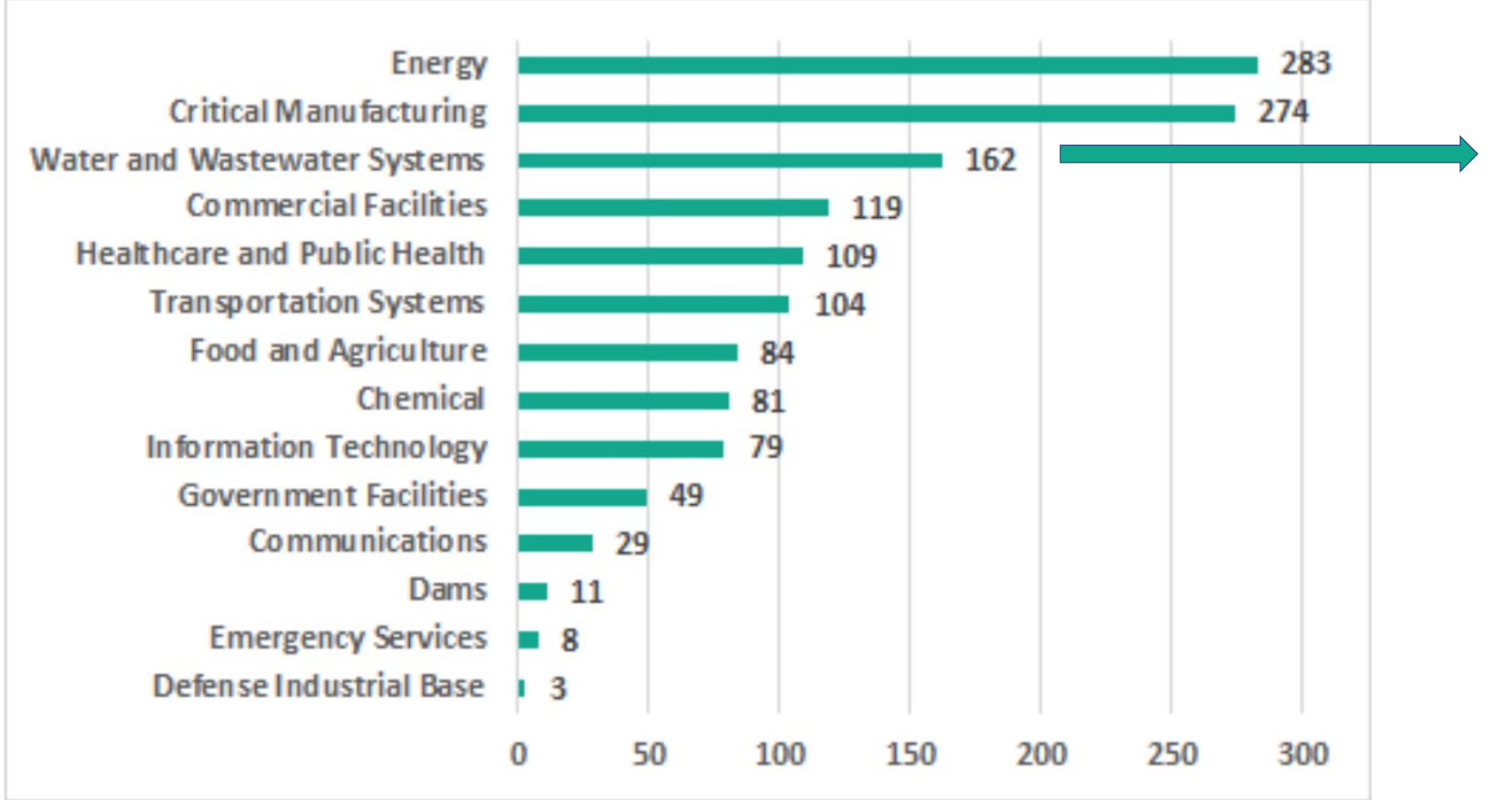


Florida Water Treatment Plant Hit with Cyber Attack

Israel Thwarts Major Coordinated Cyber Attack on Its Water Infrastructure Command and Control Systems



Critical Infrastructure sectors in the U.S.



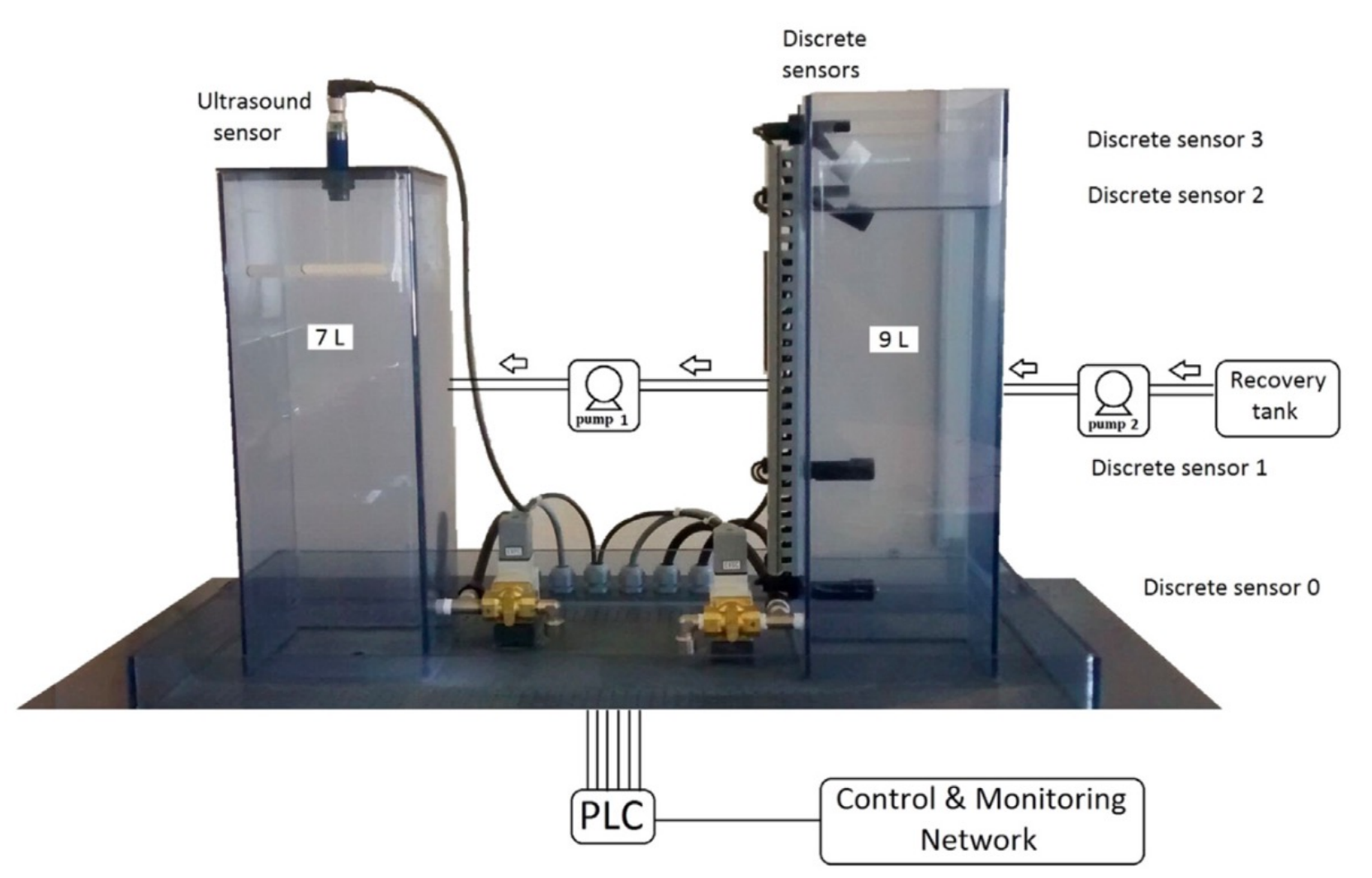
Sector	Count
Energy	283
Critical Manufacturing	274
Water and Wastewater Systems	162
Commercial Facilities	119
Healthcare and Public Health	109
Transportation Systems	104
Food and Agriculture	84
Chemical	81
Information Technology	79
Government Facilities	49
Communications	29
Dams	11
Emergency Services	8
Defense Industrial Base	3

Water Systems has 3rd number of vulnerable products used (ICS-CERT Vulnerabilities published in 2019)

CBS is related to many aspects, not just physical system, but also ethical issues, safety issues, economic issues, and so on.

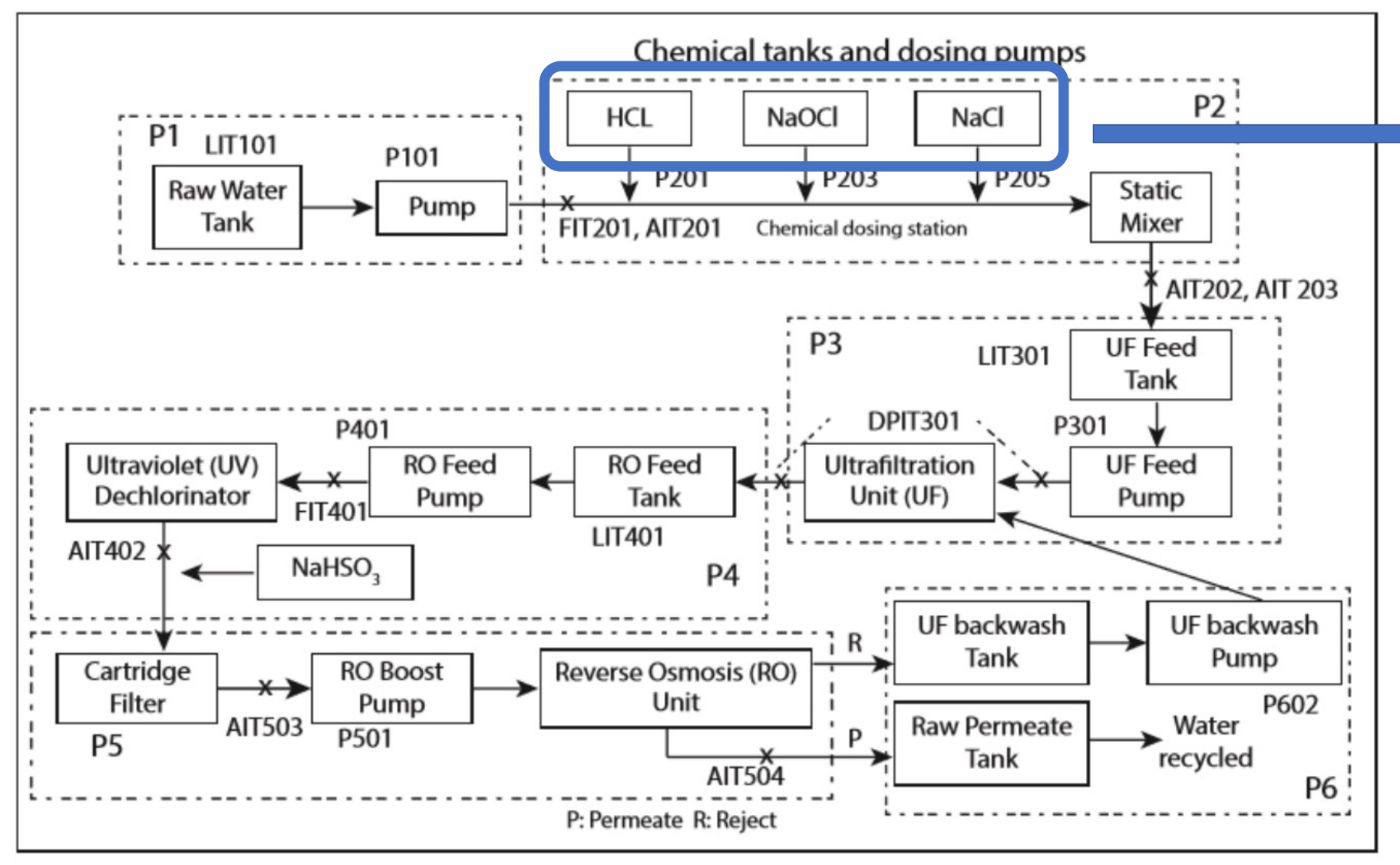
3. Case Studies

Water Distribution System – Attack Intentionality Detection



1. Two tanks with 1 ultrasound depth sensor, 4 discrete sensors, and 2 pumps
2. Contains 15 different situations related to 5 operational scenarios.
3. Labeled attack scenario with intentional and unintentional cause

Water Treatment System – Data Poisoning



Three chemical components for water quality

1. The first 7 days under normal operation and 4 days with attack scenarios
2. 4 types of attacks based on attack points in each stage

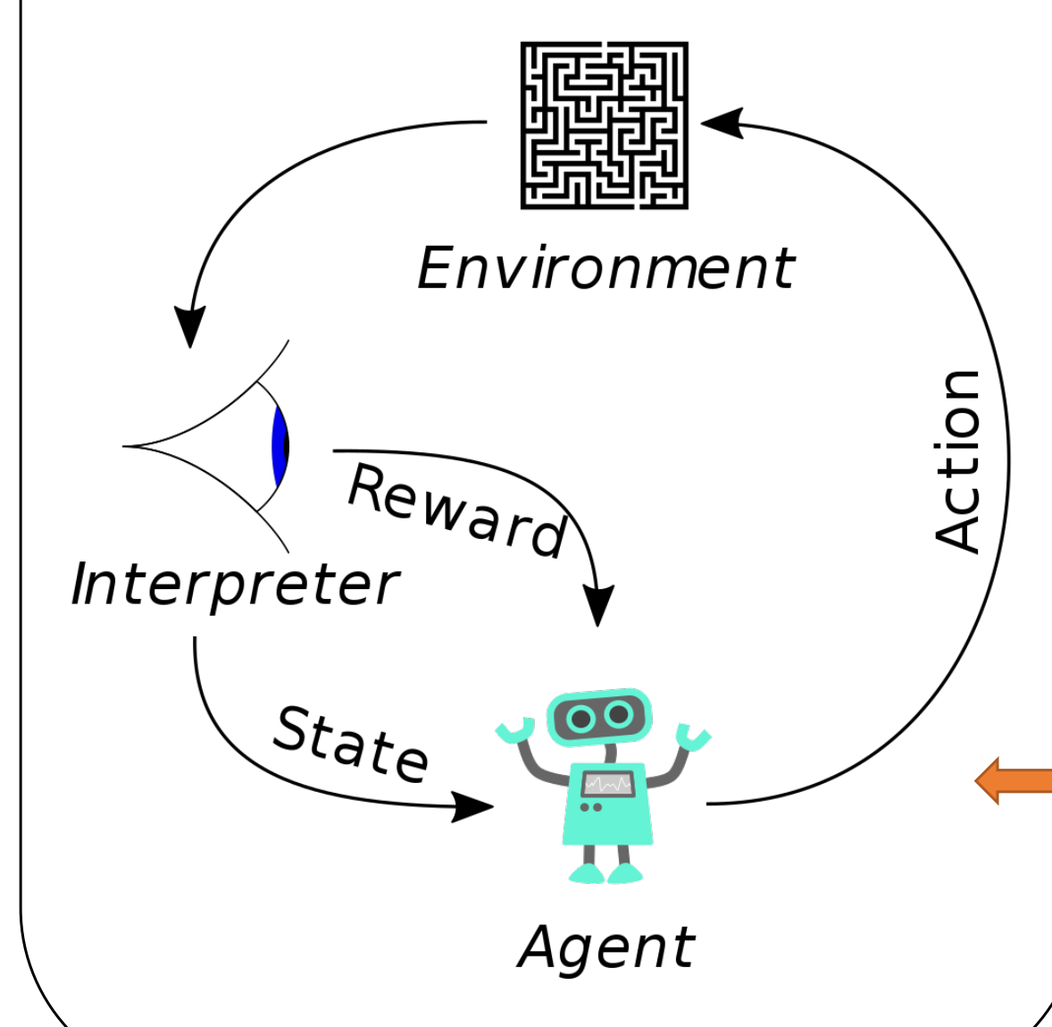
4. Model and Pipeline

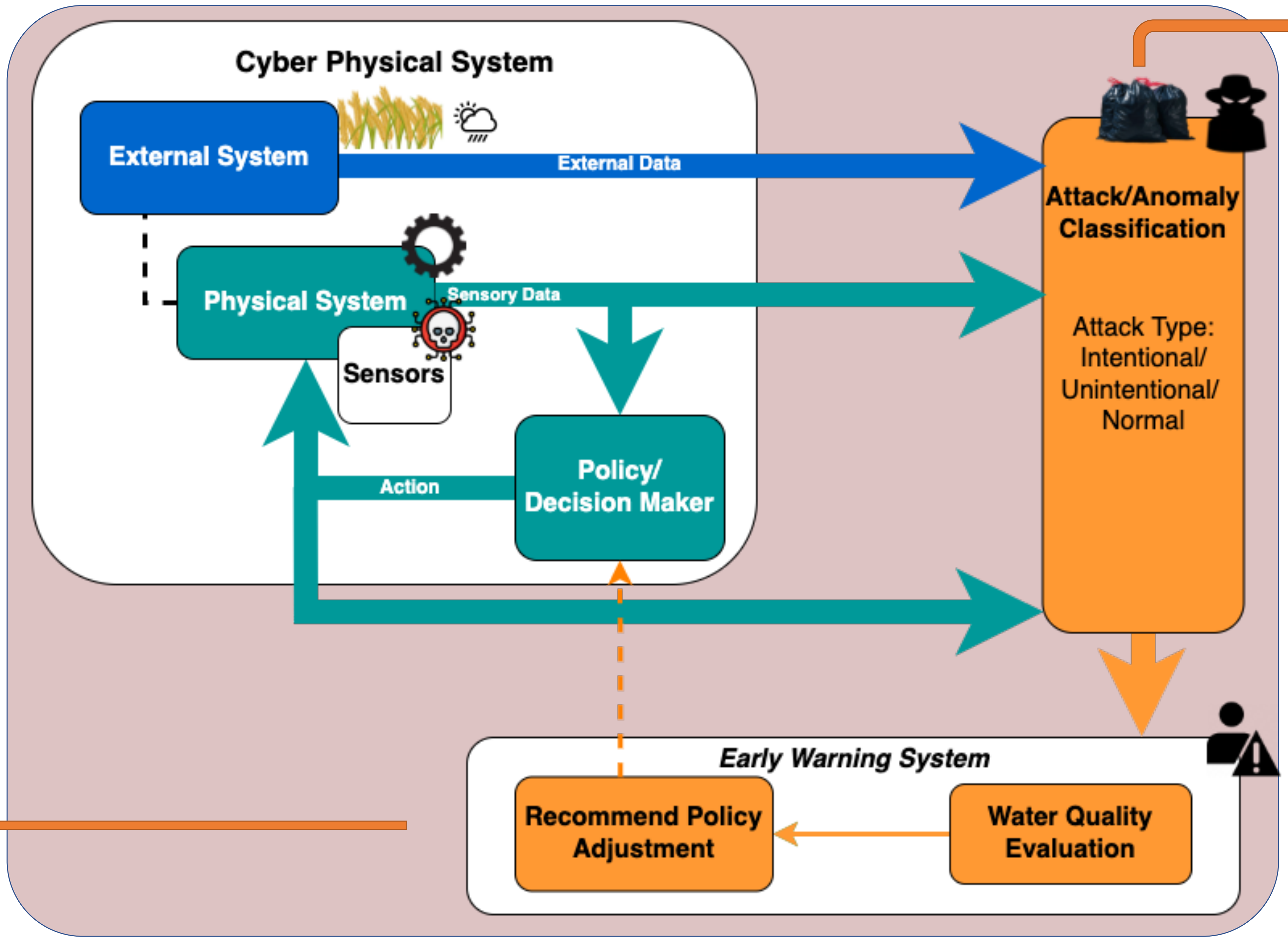
python

TensorFlow

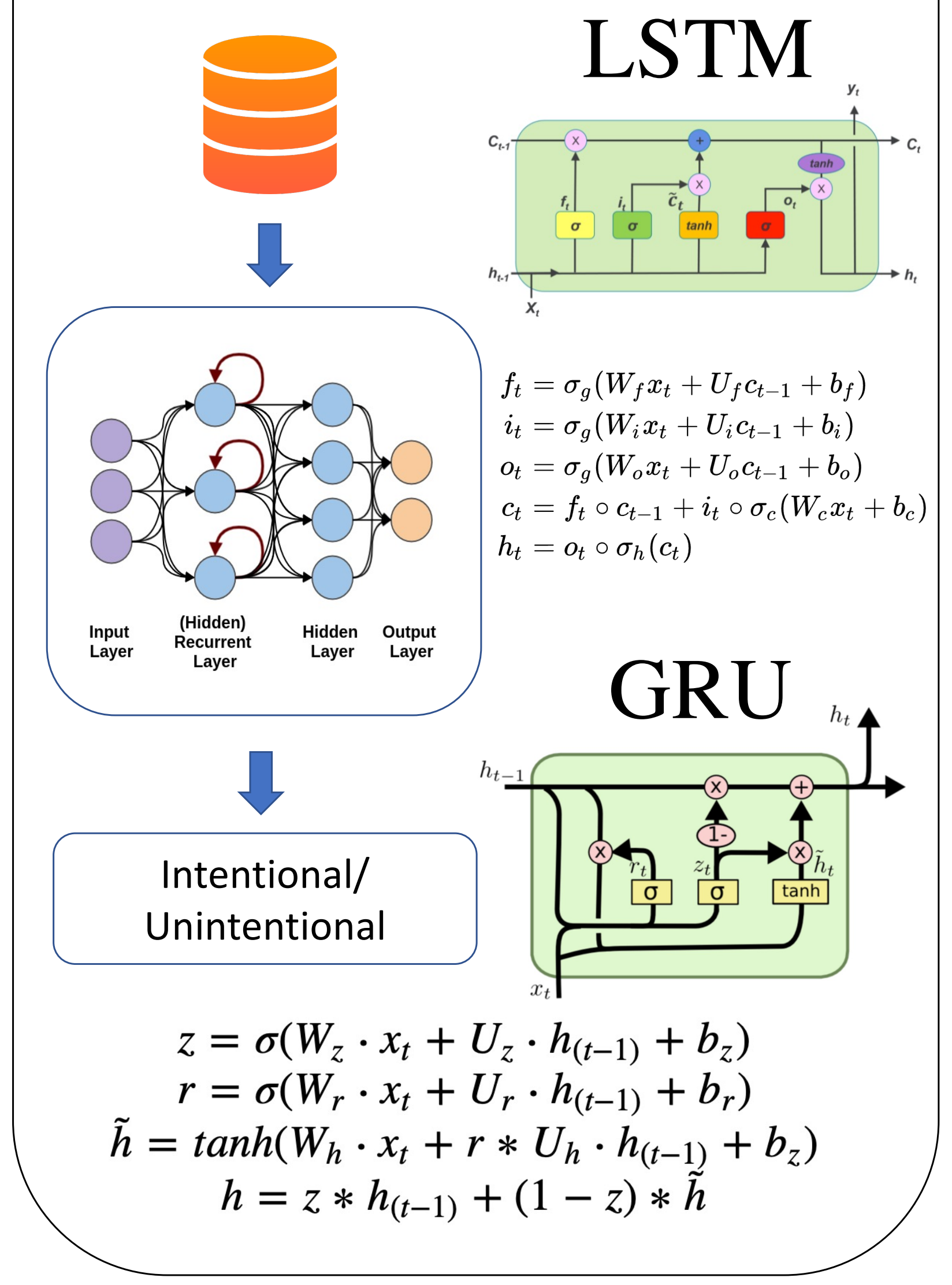
Reinforcement Learning

What is a good action to adjust the policy?





Attack/ Anomaly Classification

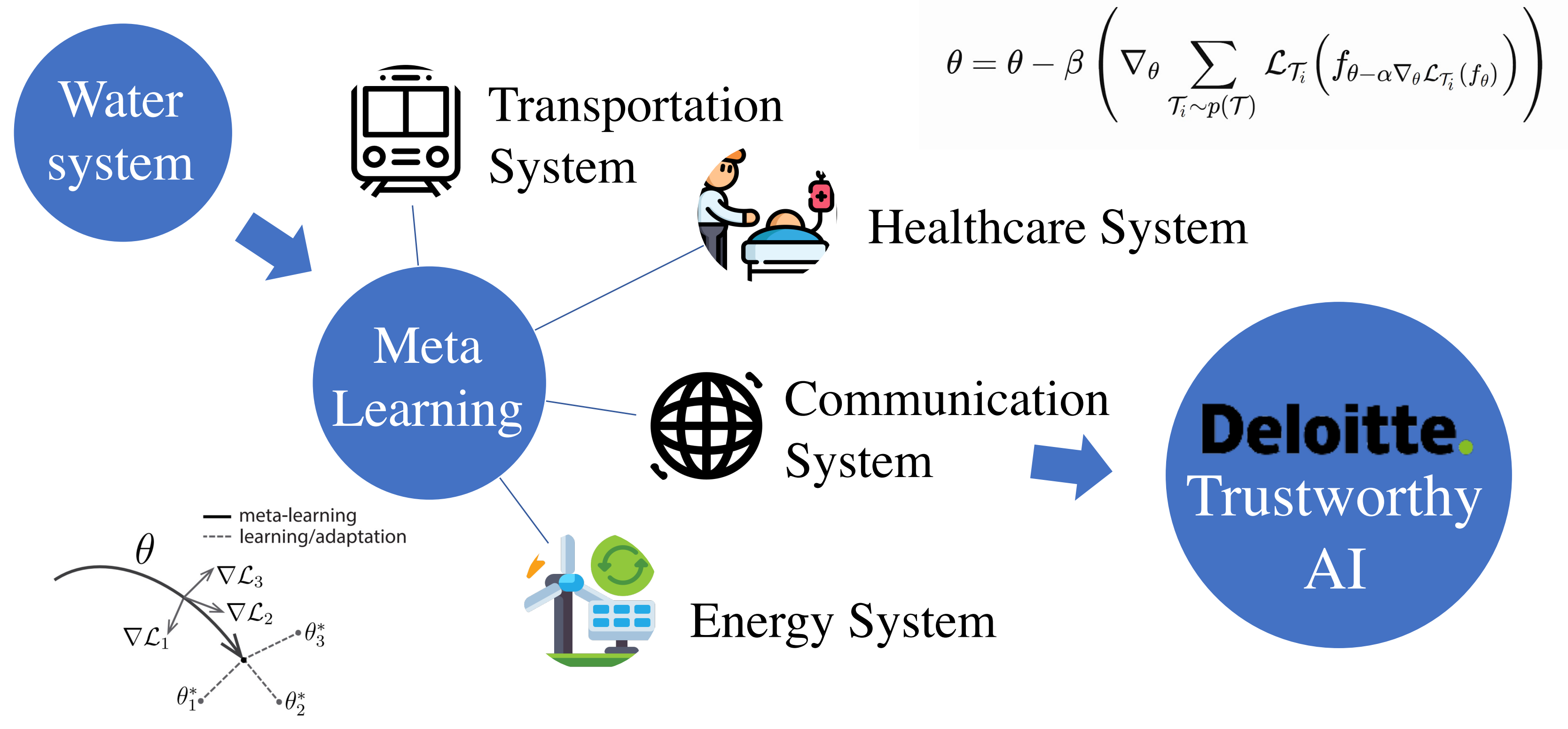


$$
\begin{aligned}
f_t &= \sigma_g(W_f x_t + U_f c_{t-1} + b_f)\\
i_t &= \sigma_g(W_i x_t + U_i c_{t-1} + b_i)\\
o_t &= \sigma_g(W_o x_t + U_o c_{t-1} + b_o)\\
c_t &= f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + b_c)\\
h_t &= o_t \circ \sigma_h(c_t)
\end{aligned}
$$

$$
\begin{aligned}
z &= \sigma(W_z \cdot x_t + U_z \cdot h_{(t-1)} + b_z)\\
r &= \sigma(W_r \cdot x_t + U_r \cdot h_{(t-1)} + b_r)\\
\tilde{h} &= tanh(W_h \cdot x_t + r * U_h \cdot h_{(t-1)} + b_z)\\
h &= z * h_{(t-1)} + (1-z) * \tilde{h}
\end{aligned}
$$

5. Future Work

$$\theta = \theta - \beta\left(\nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}\left(f_{\theta - \alpha\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)}\right)\right)$$



6. References

- Goh, Jonathan, et al. "A dataset to support research in the design of secure water treatment systems." Proceedings of the International Conference on Critical Information Infrastructures Security. Springer, Cham, 2016.
- Laso, Pedro Merino, David Brosset, and John Puentes. "Dataset of anomalies and malicious acts in a cyber-physical subsystem." Data in brief 14 (2017): pp. 186-191.
- Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks." Proceedings of the International Conference on Machine Learning. PMLR, 2017

Presented at the 35th International Florida Artificial Intelligence Research Conference (FLAIRS) – Jensen Beach, FL