



Explainable AI as Interface between Users and Autonomic Systems: the Case of the Smart Home

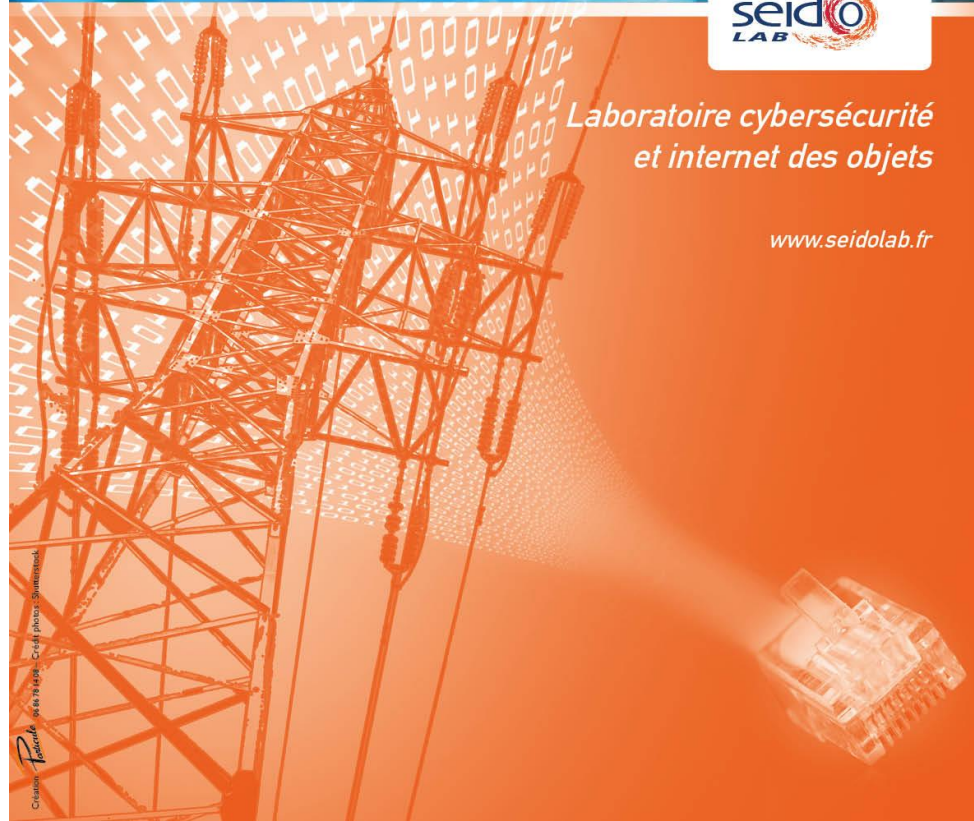
HOUZE Etienne
DIACONESCU Ada
DESSALLES Jean-Louis
MENGA David

WORKSHOP SEIDO – 29 janvier 2018



Laboratoire cybersécurité
et internet des objets

www.seidolab.fr



Citation: P. Radocha, 06/10/18 Edge - Credit photos: Shutterstock

Un domaine prometteur



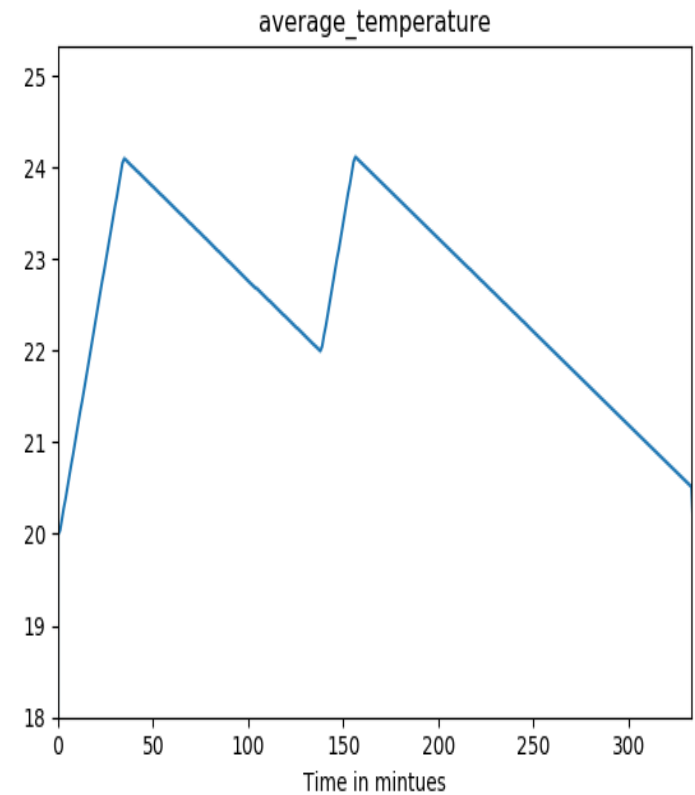
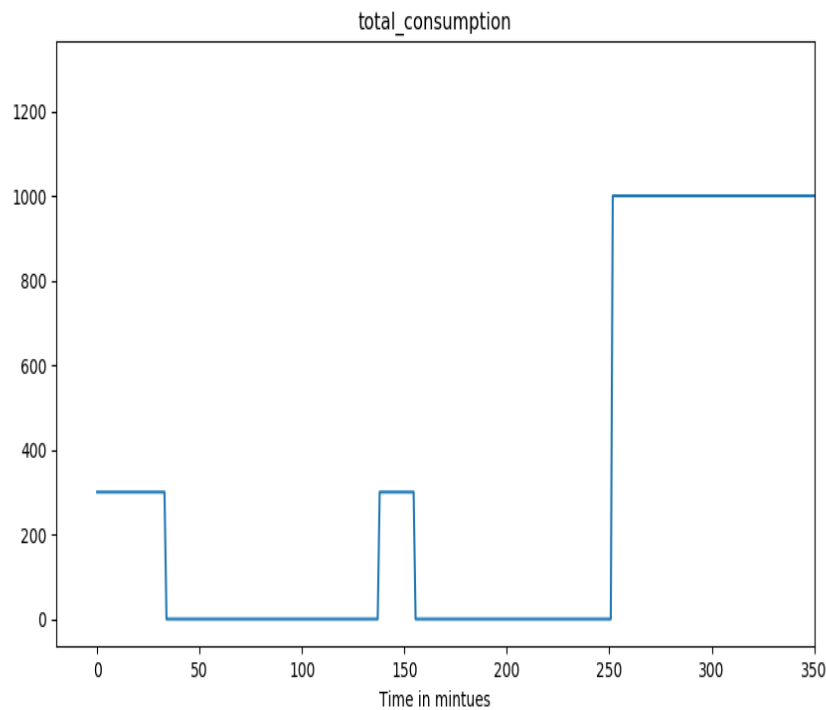
EDF Soweé©

Amazon Echo©

Philips Hue©

Des freins à l'explicabilité

- Système autonome à variables numériques



Des freins à l'explicabilité

- Système autonome à variables numériques
- Langage humain “flou” et dynamique

Des freins à l'explicabilité

- Système autonome à variables numériques
- Langage humain “flou” et dynamique
 - “Il fait froid” ?

Des freins à l'explicabilité

- Système autonome à variables numériques
- Langage humain “flou” et dynamique
 - “Il fait froid” ?
- Système générique et adaptable

Approche proposée

- Utiliser des prédicats
- Abandon des règles fixées
- Décentraliser le raisonnement

Utiliser des prédicats

- Prédicat : Fonction booléenne s'appliquant sur un sous-ensemble des variables mesurées par le système.



Utiliser des prédicats

- Prédicat : Fonction booléenne s'appliquant sur un sous-ensemble des variables mesurées par le système.
- Toute observable qualifiable par l'utilisateur est un prédicat, associé à une *désirabilité*.

Utiliser des prédicats

- Prédicat : Fonction booléenne s'appliquant sur un sous-ensemble des variables mesurées par le système.
- Toute observable qualifiable par l'utilisateur est un prédicat, associé à une *désirabilité*.
 - *Confort(Salon), 15*

Utiliser des prédicats

- Prédicat : Fonction booléenne s'appliquant sur un sous-ensemble des variables mesurées par le système.
- Toute observable qualifiable par l'utilisateur est un prédicat, associé à une *désirabilité*.
 - *Confort(Salon), 15*
 - *Froid(Salon), -20*

Utiliser des prédicats

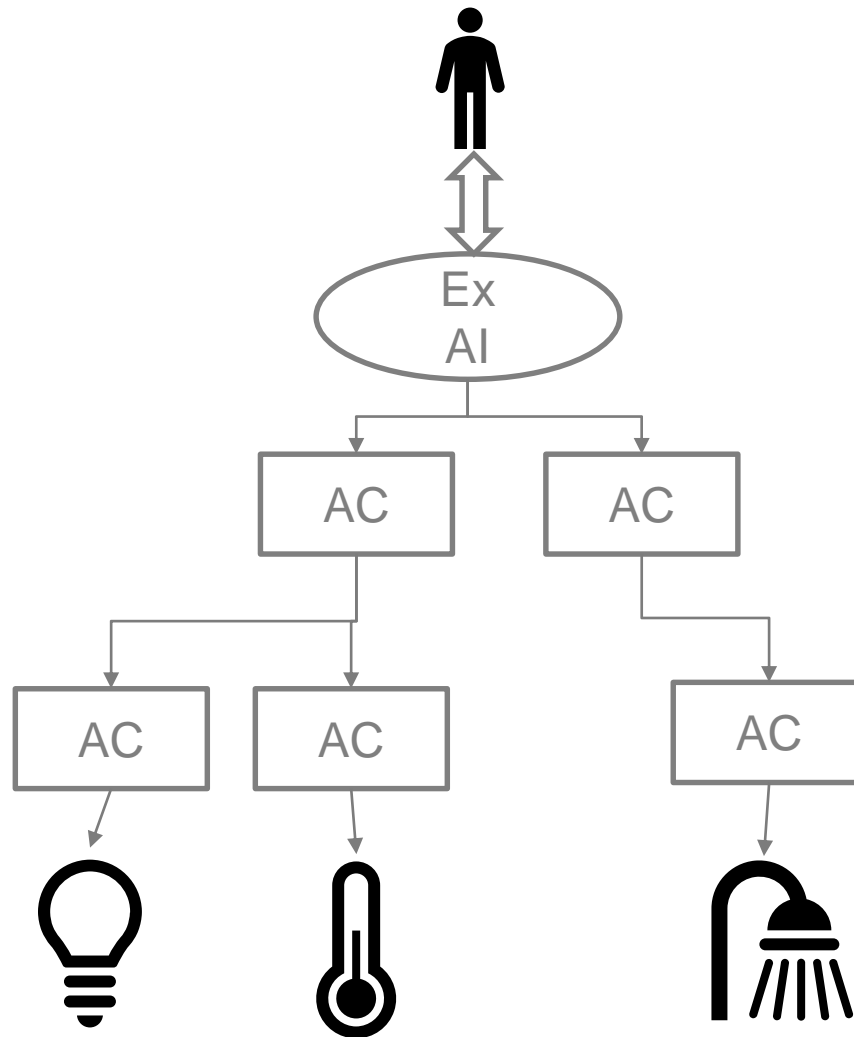
- Prédicat : Fonction booléenne s'appliquant sur un sous-ensemble des variables mesurées par le système.
- Toute observable qualifiable par l'utilisateur est un prédicat, associé à une *désirabilité*.
 - *Confort(Salon), 15*
 - *Froid(Salon), -20*
 - *BlackOut(Maison), -50*

Utiliser des prédicats

- Prédicat : Fonction booléenne s'appliquant sur un sous-ensemble des variables mesurées par le système.
- Toute observable qualifiable par l'utilisateur est un prédicat, associé à une *désirabilité*.
 - *Confort(Salon), 15*
 - *Froid(Salon), -20*
 - *BlackOut(Maison), -50*
- En l'absence de règles, explication par propagation du conflit.

Décentraliser le raisonnement

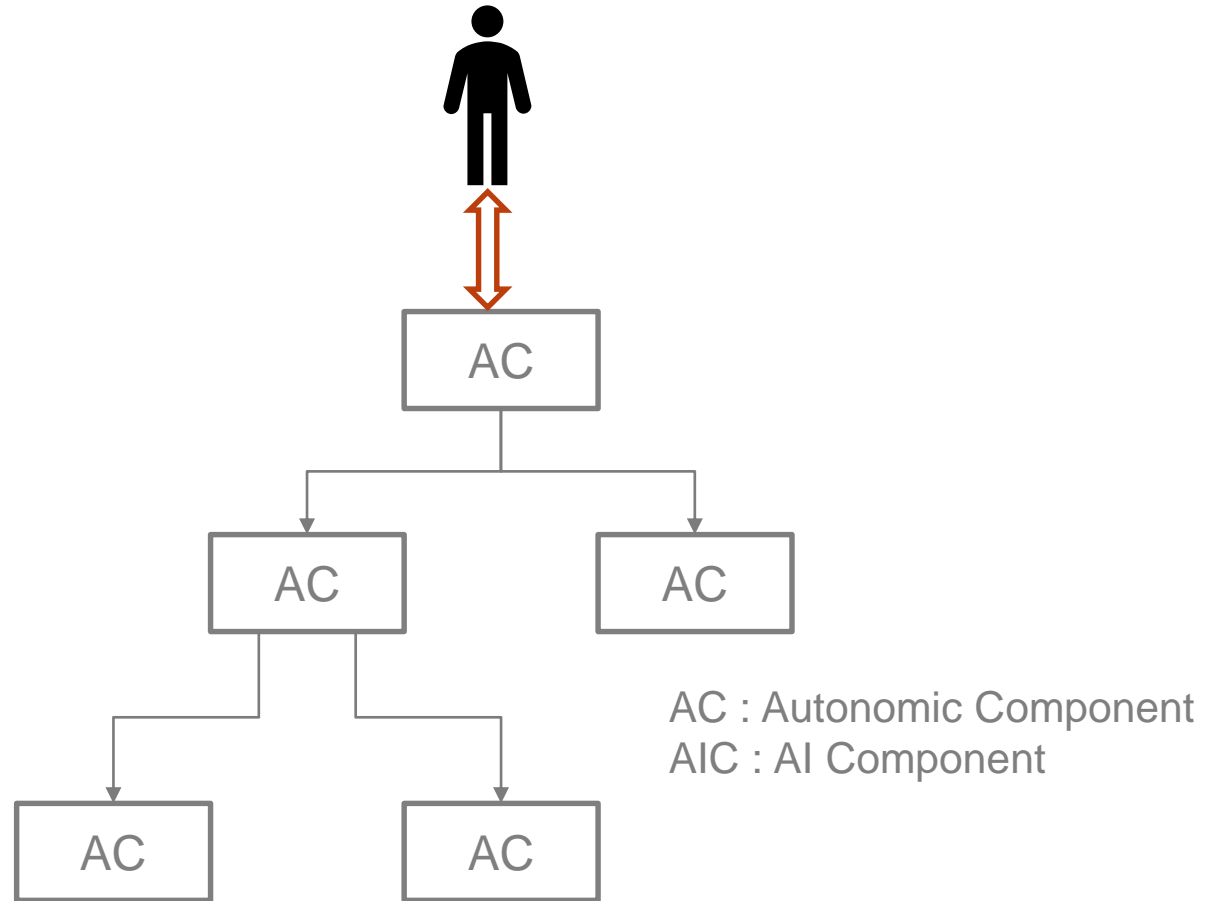
Un système centralisé



AC : Autonomic Component

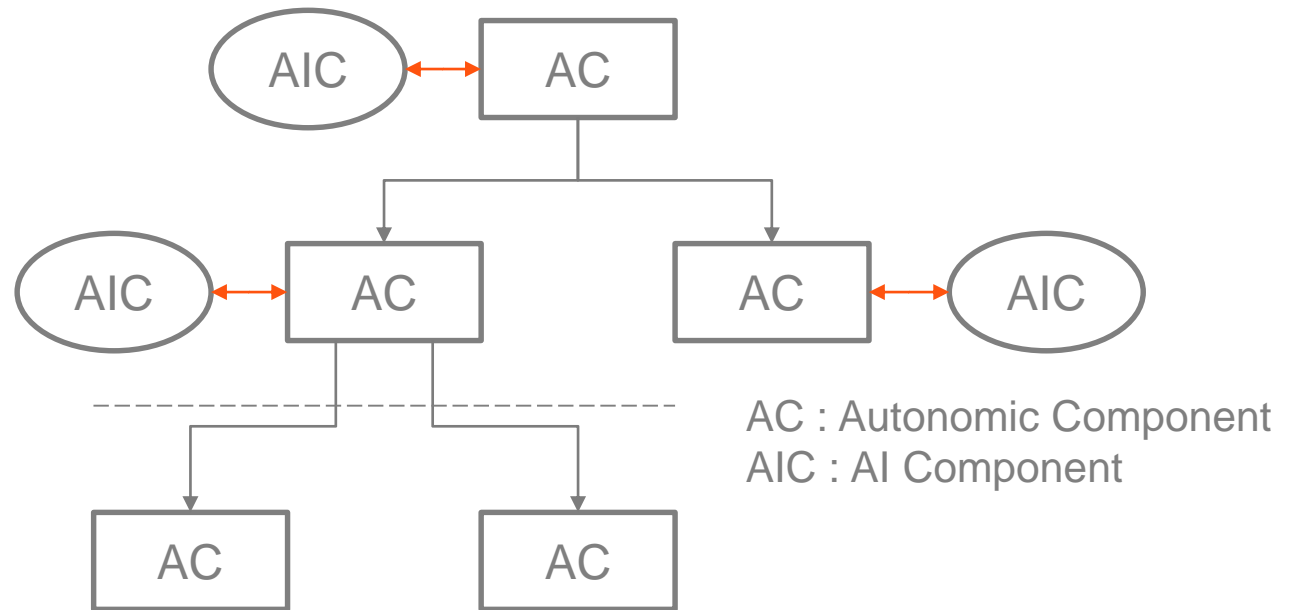
Décentraliser le raisonnement

Notre approche



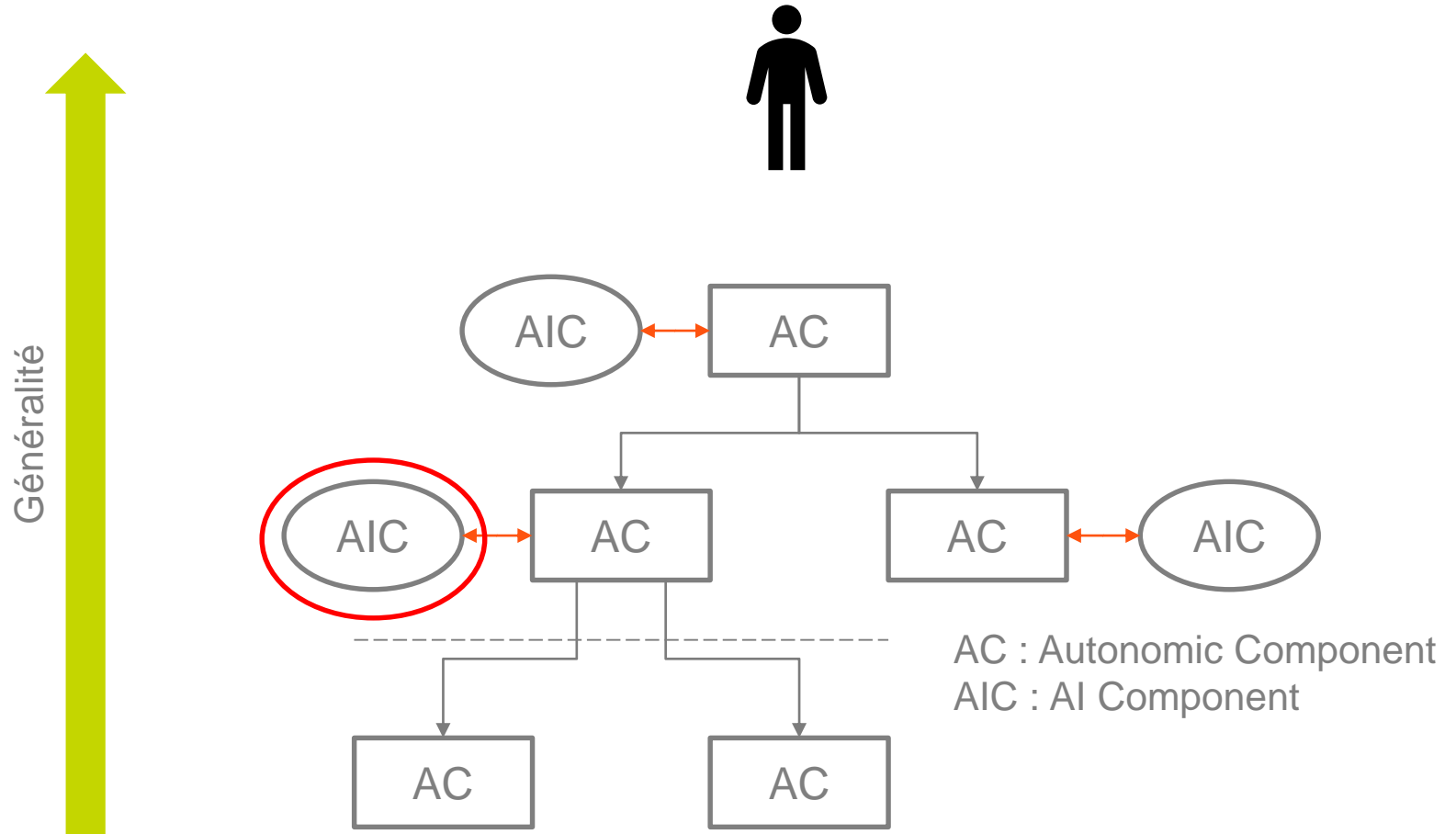
Décentraliser le raisonnement

Notre approche



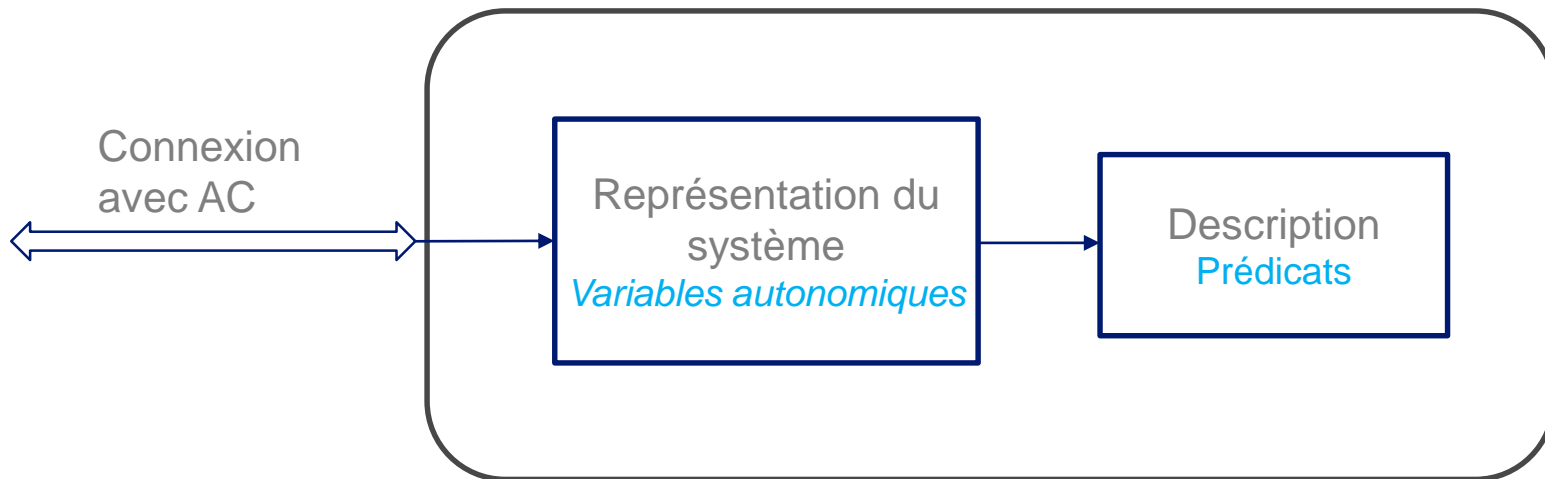
Décentraliser le raisonnement

Notre approche



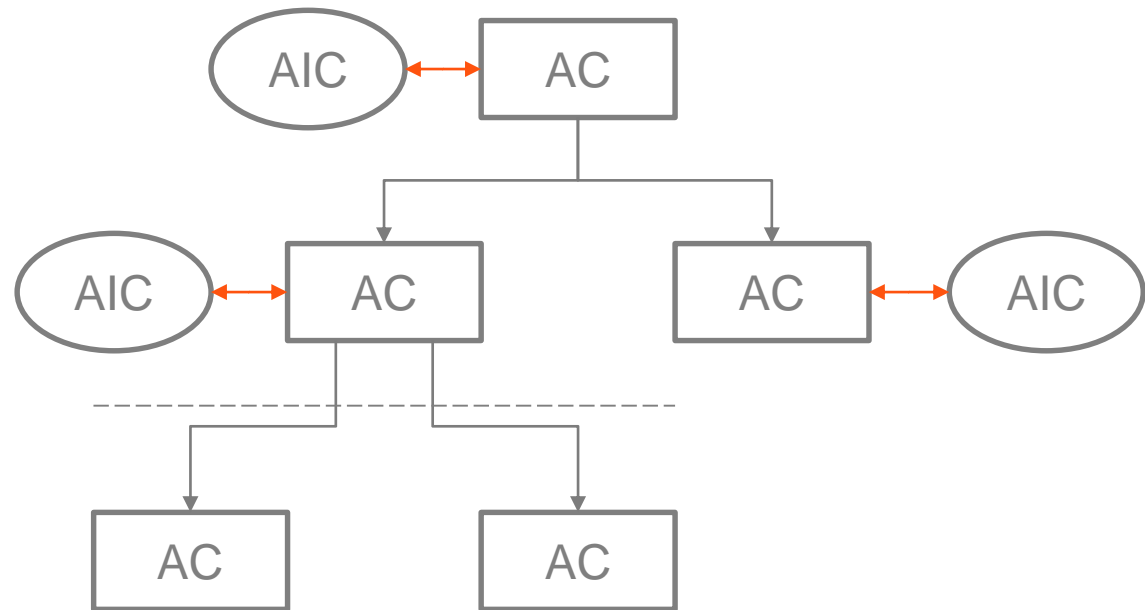
Décentraliser le raisonnement

AI Component



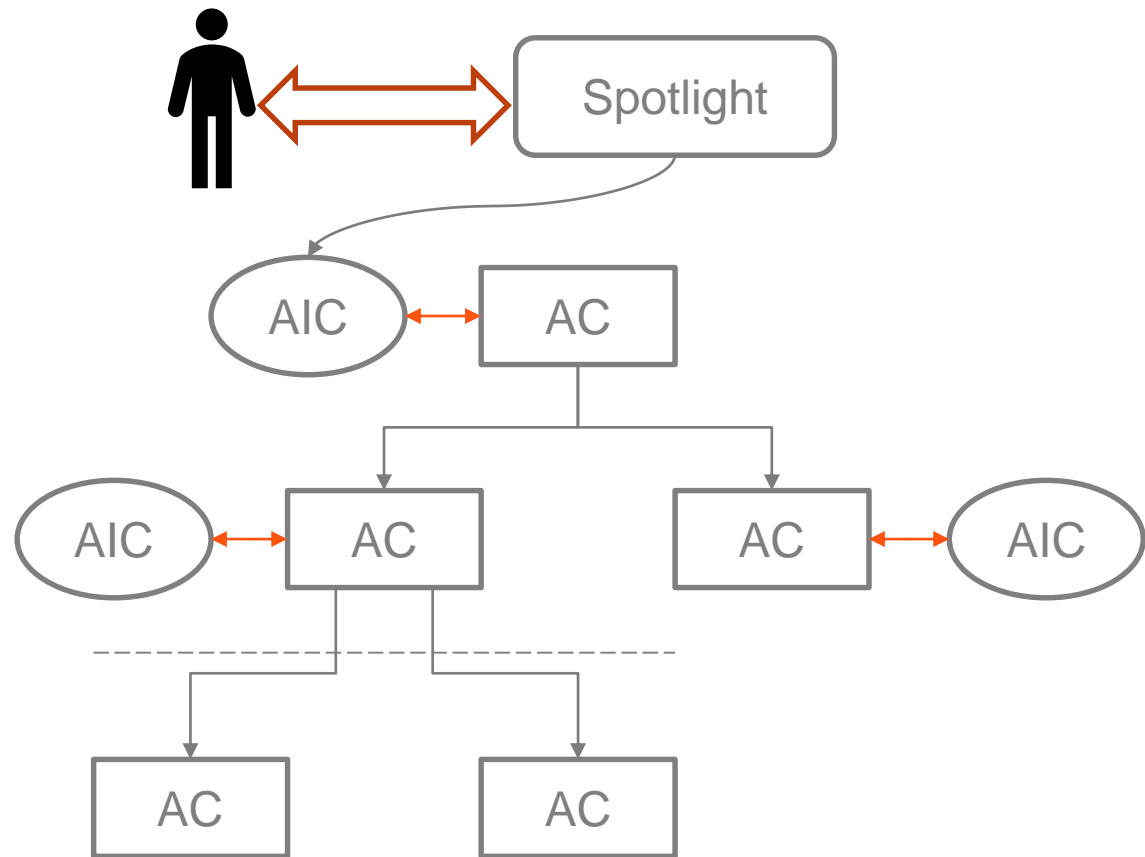
Décentraliser le raisonnement

Ajout du “spotlight”



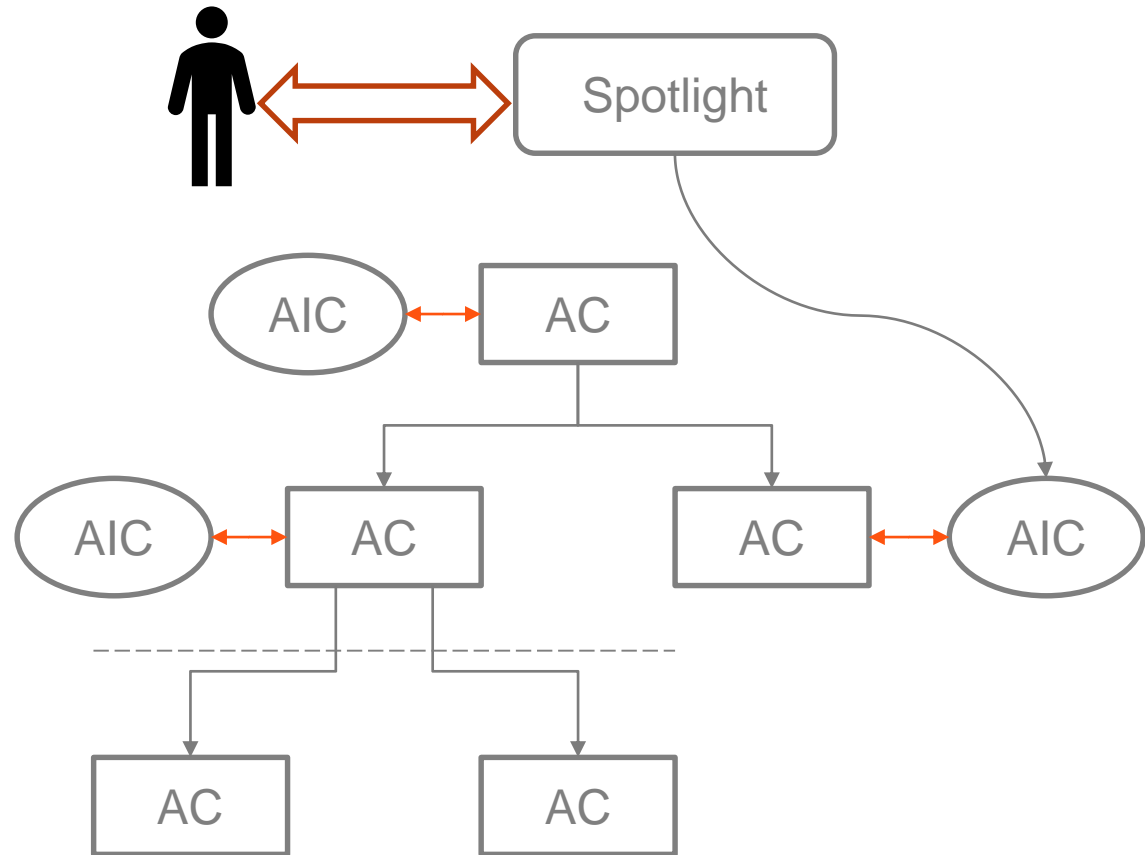
Décentraliser le raisonnement

Ajout du “spotlight”



Décentraliser le raisonnement

Ajout du “spotlight”



Scénarios d'exemple

- Objectif : Expliquer sans règles

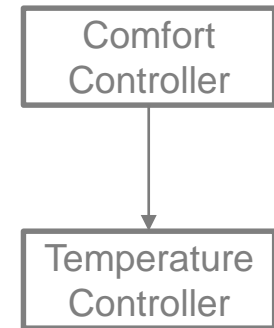
Scénarios d'exemple

- Objectif : Expliquer sans règles
- Situation initiale :
 - Un contrôleur de confort

Comfort
Controller

Scénarios d'exemple

- Objectif : Expliquer sans règles
- Situation initiale :
 - Un contrôleur de confort
 - Un contrôleur de température

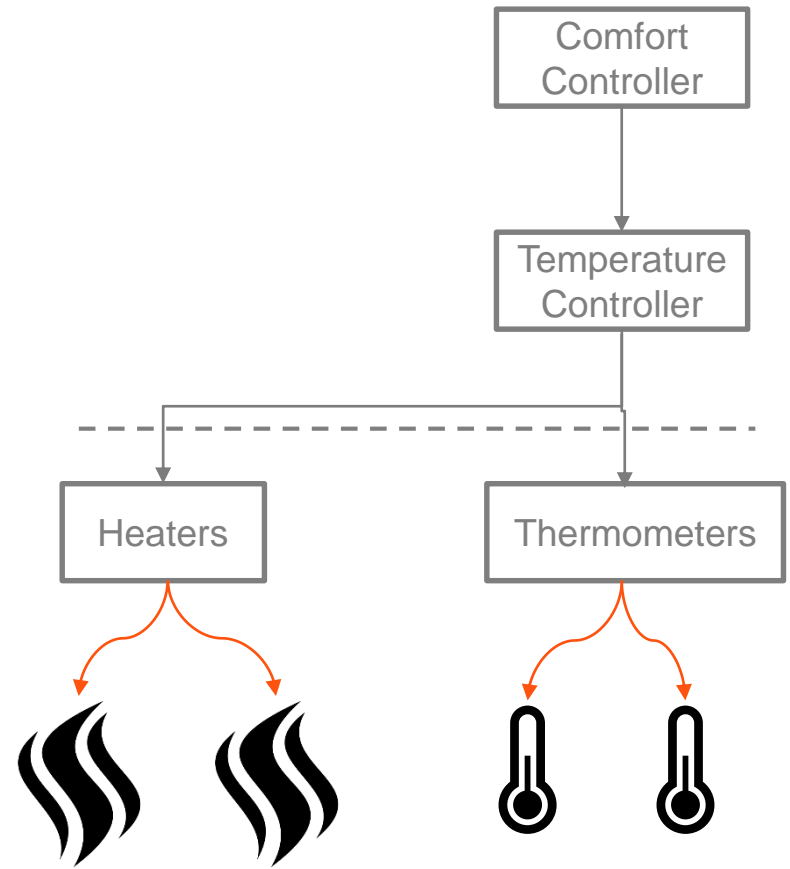


Scénarios d'exemple

■ Objectif : Expliquer sans règles

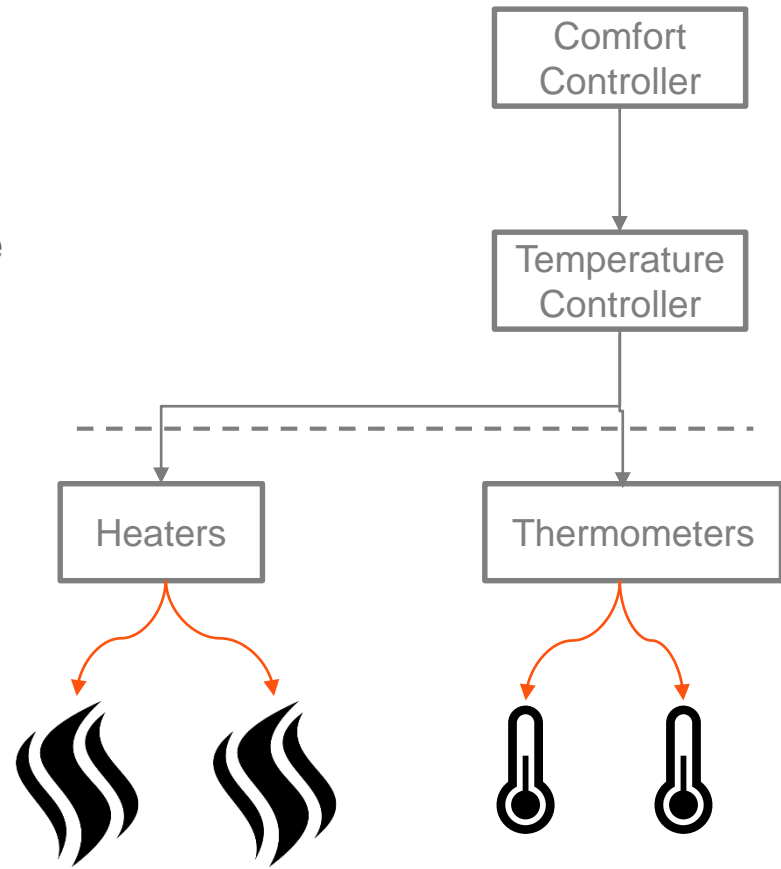
■ Situation initiale :

- Un contrôleur de confort
- Un contrôleur de température
- Un thermostat et un thermomètre



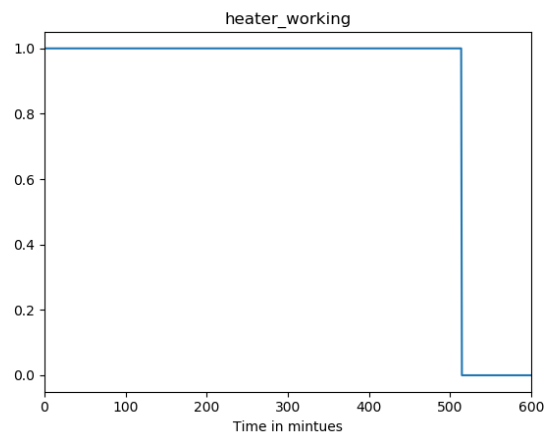
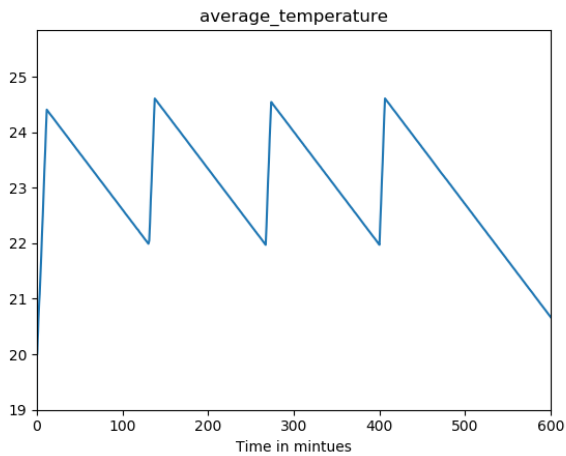
Scénarios d'exemple

- Objectif : Expliquer sans règles
- Scénario 1 : Conflit
 - Panne dans le thermostat
 - Anomalie de température détectée
 - Questionnement de l'utilisateur
 - Explication par le conflit antérieur



Scénarios d'exemple

- Objectif : Expliquer sans règles
- Scénario 1 : Conflit

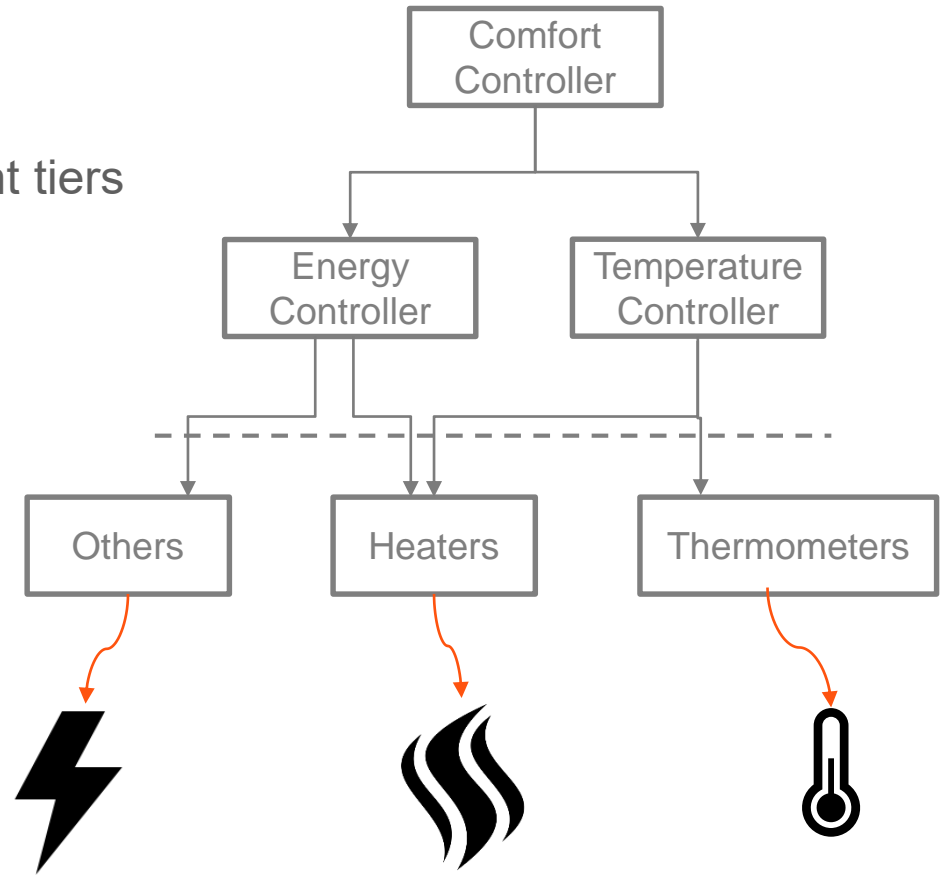


Scénarios d'exemple

- Objectif : Expliquer sans règles

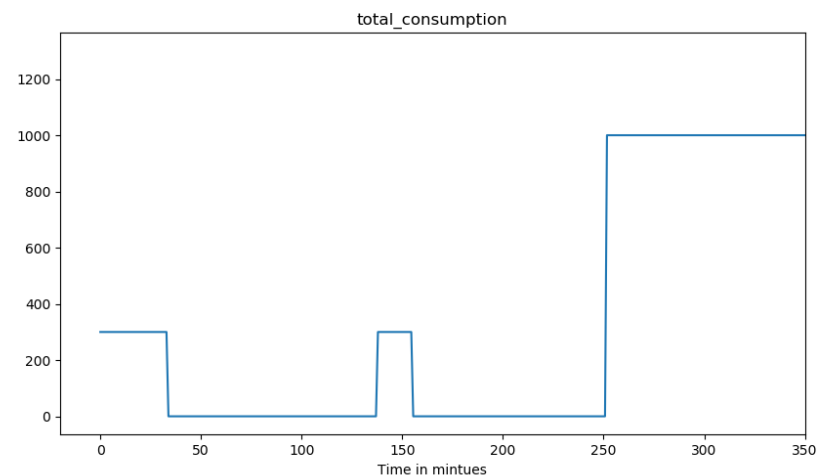
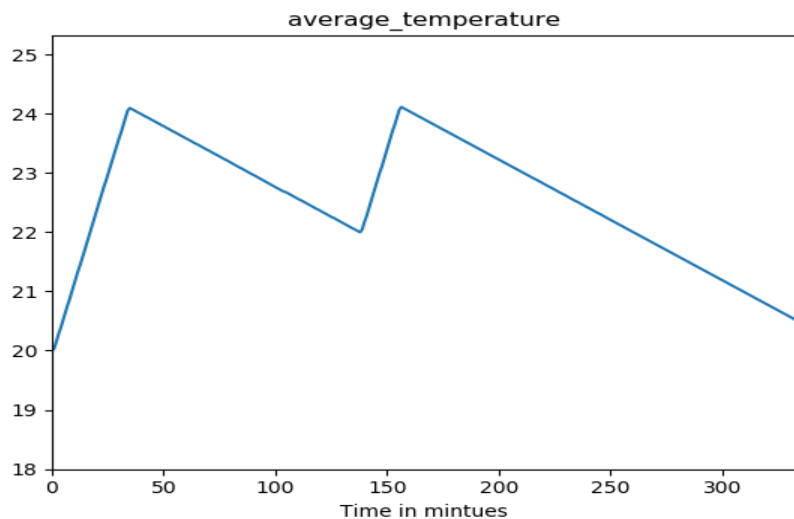
- Scénario 2 : Simulation

- Consommation d'un composant tiers
- Arrêt du chauffage
- Conflit
- Simulation
- Explication



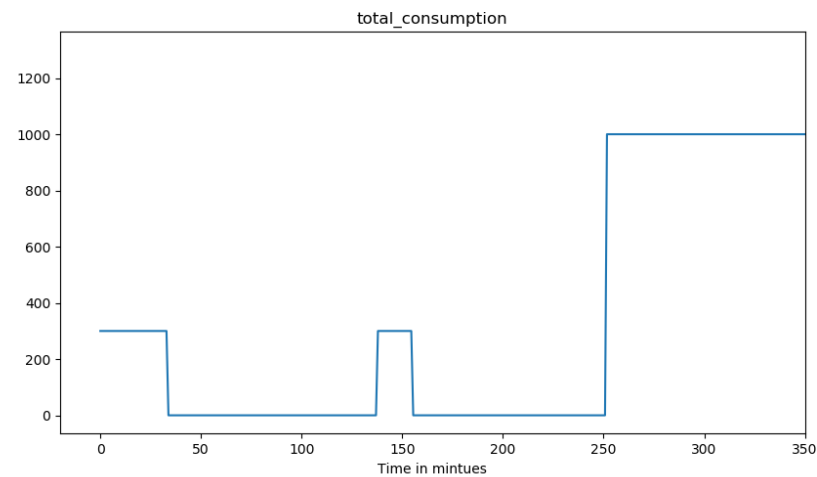
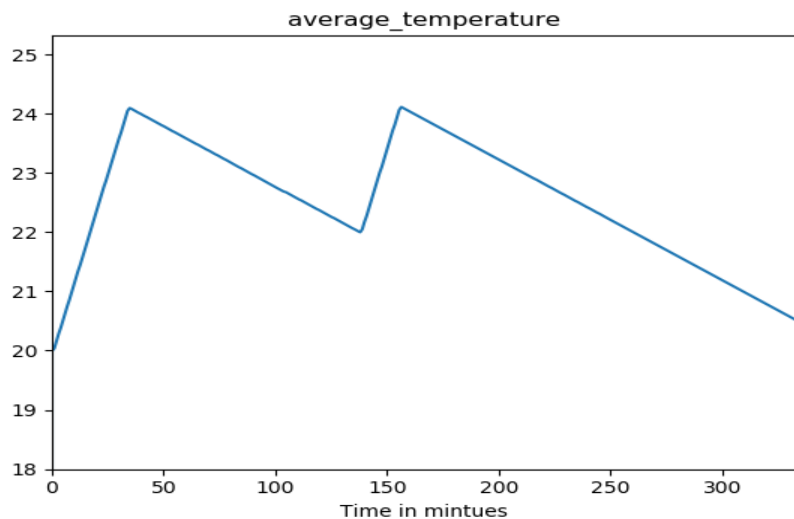
Scénarios d'exemple

- Objectif : Expliquer sans règles
- Scénario 2 : Simulation



Scénarios d'exemple

- Objectif : Expliquer sans règles
- Scénario 2 : Simulation



Scénarios d'exemple

- Objectif : Expliquer sans règles
- Scénario 2 : Simulation

```
[QUESTION] why cold ?  
[CONFLICT] not Comfort is the cause of cold  
[QUESTION] Are you satisfied ?  
[USER_INPUT] No  
[CONFLICT] conflict found !  
Cold because if NotCold then BlackOut
```

Conclusion

- Problématiques
 - Expliquer un système autonome
 - Adaptabilité au système
 - Gestion de l'imprévu
- Réponses proposées :
 - Interface variables autonomiques / prédicats
 - Décentralisation du raisonnement
 - Absence de règles

Perspectives

- Progrès actuel
 - Définitions des principes guides
 - Démonstration de cas basiques
- Futur :
 - Abandon progressif des seuils
 - Application à plus grande échelle