CHALMERS UNIVERSITY OF TECHNOLOGY | UNIVERSITY OF GOTHENBURG

Software-Enabled Solutions for Human-on-the-Loop Autonomous Systems

Rebekka Wohlrab


S02-A3


ABB

Humans as bystanders

Humans as maintainers

Human as input-givers

Human as collaborators

Humans as the ones that do what a system can't do

R.D. Alexander, N.J. Herbert, & T. Kelly. The role of the human in an Autonomous System. IET Conference Publications. 2009. 1 - 6. 10.1049/cp.2009.1536.

Challenges of human-machine collaboration

- Humans and machines speak different languages
- Complexity of the problem space with many relevant qualities (safety, security, performance, cost, reliability, ...)
- Uncertainty
- Trust
- ...

Baraglia, J., Cakmak, M., Nagai, Y., Rao, R. P., & Asada, M. (2017). Efficient human-robot collaboration: when should a robot take initiative?. *The International Journal of Robotics Research*, *36*(5-7), 563-579.
Xiong, W., Fan, H., Ma, L., & Wang, C. (2022). Challenges of human—machine collaboration in risky decision-making. *Frontiers of Engineering Management*, *9*(1), 89-103.

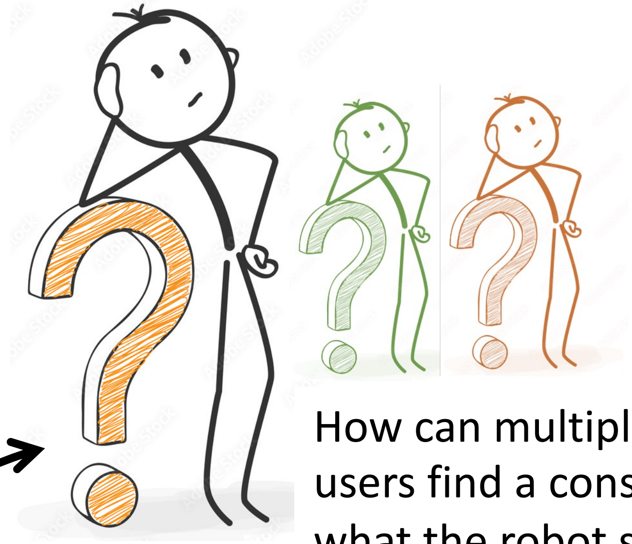
Human-on-the-loop autonomous systems

What does the human do?

What does the robot do?

What does the human want the robot to do?

What does the robot want the human to do?



How can multiple human users find a consensus on what the robot should do?

Human-on-the-loop autonomous systems

Monitoring [4]

Tradeoff Explainability [1]

What does the human do?

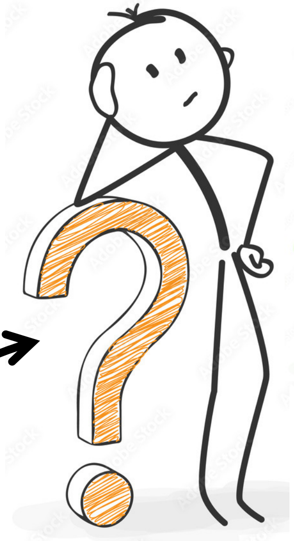
What does the robot do?

What does the human want the robot to do?

Preference elicitation [2, 3]

This Photo by Unknown Author is licensed under CC BY-SA-NC

What does the robot want the human to do?



Negotiation [2]

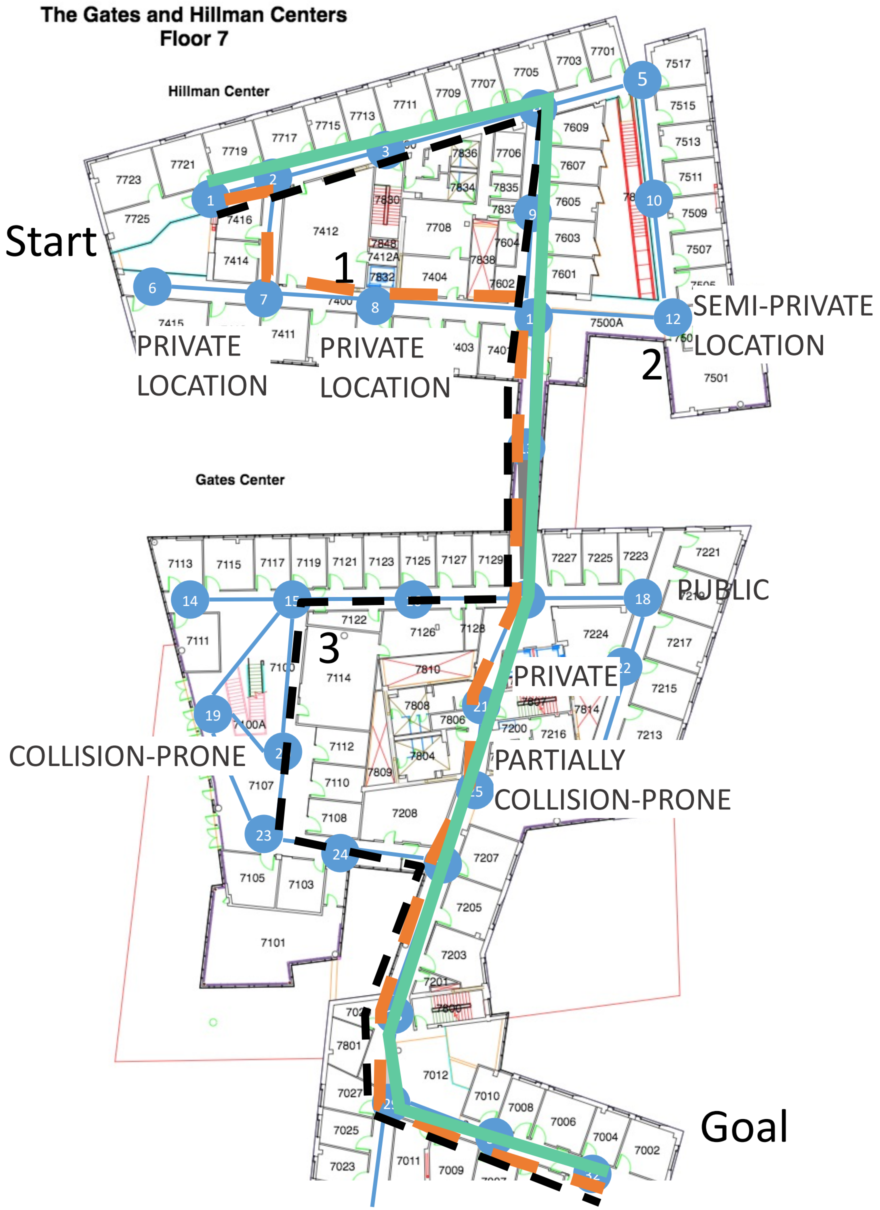
How can multiple human users find a consensus on what the robot should do?

[1] Wohlrab, R., Cámara, J., Garlan, D., & Schmerl, B. (2023). Explaining quality attribute tradeoffs in automated planning for self-adaptive systems. *Journal of Systems and Software.*

[2] Wohlrab, R., & Garlan, D. (2022). A negotiation support system for defining utility functions for multi-stakeholder self-adaptive systems. *Requirements Engineering*.

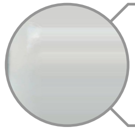
[3] Wohlrab, R., Meira-Góes, R., & Vierhauser, M. (2022). Run-time adaptation of quality attributes for automated planning. In *Proceedings of the 17th Symposium on Software Engineering for Adaptive and Self-Managing Systems* (SEAMS).

[4] Vierhauser, M., Wohlrab, R., Stadler, M., & Cleland-Huang, J. (2023). AMon: A domain-specific language and framework for adaptive monitoring of Cyber–Physical Systems. *Journal of Systems and Software*.



Qualities in automated mission planning

Safety
(collision avoidance)



Privacy

Efficiency (travel time)

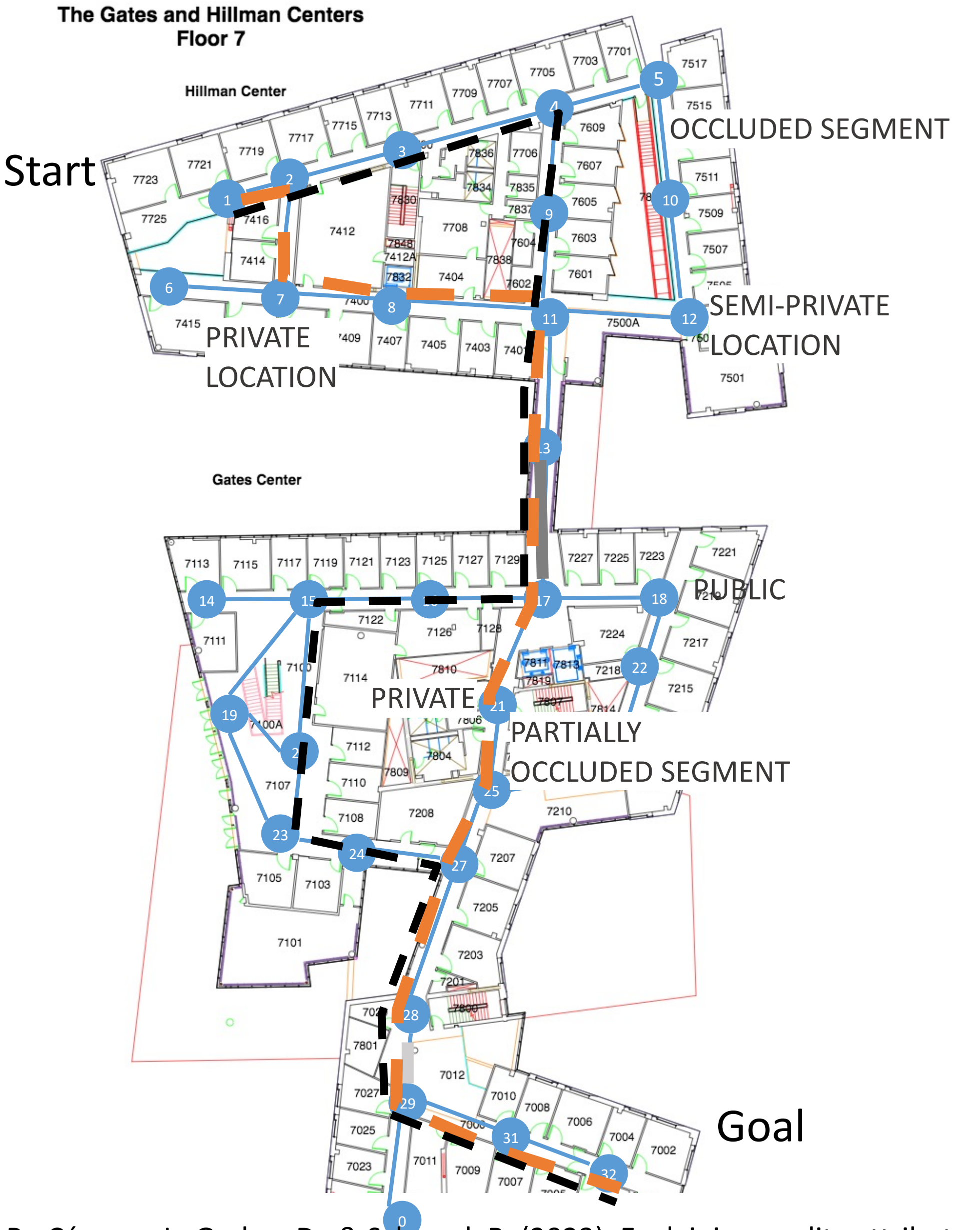
- Problems:
  - Autonomous systems often do not allow for changing priorities of objectives at runtime
  - Tradeoffs tend to be opaque

Option 1:
Travel time
Safety
Privacy

Option 2:
Travel time
Safety
Privacy

Option 3:
Travel time
Safety
Privacy

Wohlrab, R., Cámara, J., Garlan, D., & Schmerl, B. (2023). Explaining quality attribute tradeoffs in automated planning for self-adaptive systems. *Journal of Systems and Software*, *198*, 111538.



How to select a plan?

- Using a planner
- It needs an optimization function
- Need to indicate weights/priorities for different objectives
  - Safety: rather unimportant (0.1)
  - Privacy: rather unimportant (0.1)
  - Travel time: very important (0.8)

$$utility(plan) = 0.8 \cdot utility\_travel\_time(plan) + 0.1 \cdot utility\_safety(plan) + 0.1 \cdot utility\_privacy(plan)$$

- What should those weights/priorities be?
- How do they impact the generated plans?

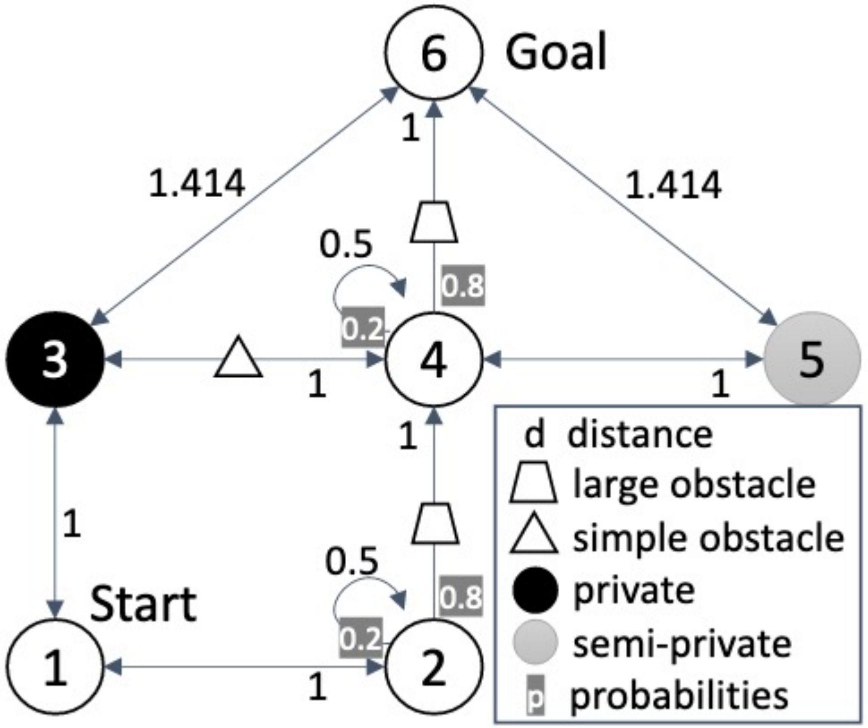
Wohlrab, R., Cámara, J., Garlan, D., & Schmerl, B. (2023). Explaining quality attribute tradeoffs in automated planning for self-adaptive systems. *Journal of Systems and Software*, *198*, 111538.

Markov Decision Process

```
mdp
formula goal = rLoc=6;
label "end" = rLoc=6 & !computeGo & barrier;

module module_1
  rLoc : [0..6] init 1;
  // moveTo
  [moveTo_LP_L2_RP_] rLoc=1 -> 0.0:(rLoc'=1) + 1.0:(rLoc'=2);
  [moveTo_LP_L4_RP_] rLoc=2 -> 0.2:(rLoc'=2) + 0.8:(rLoc'=4);
  //...
endmodule

rewards "travelTime"
  [moveTo_LP_L2_RP_] rSpeed=0 & rLoc=1 : 1.0;
  [moveTo_LP_L4_RP_] rSpeed=1 & rLoc=2 : 0.8;
endrewards
```



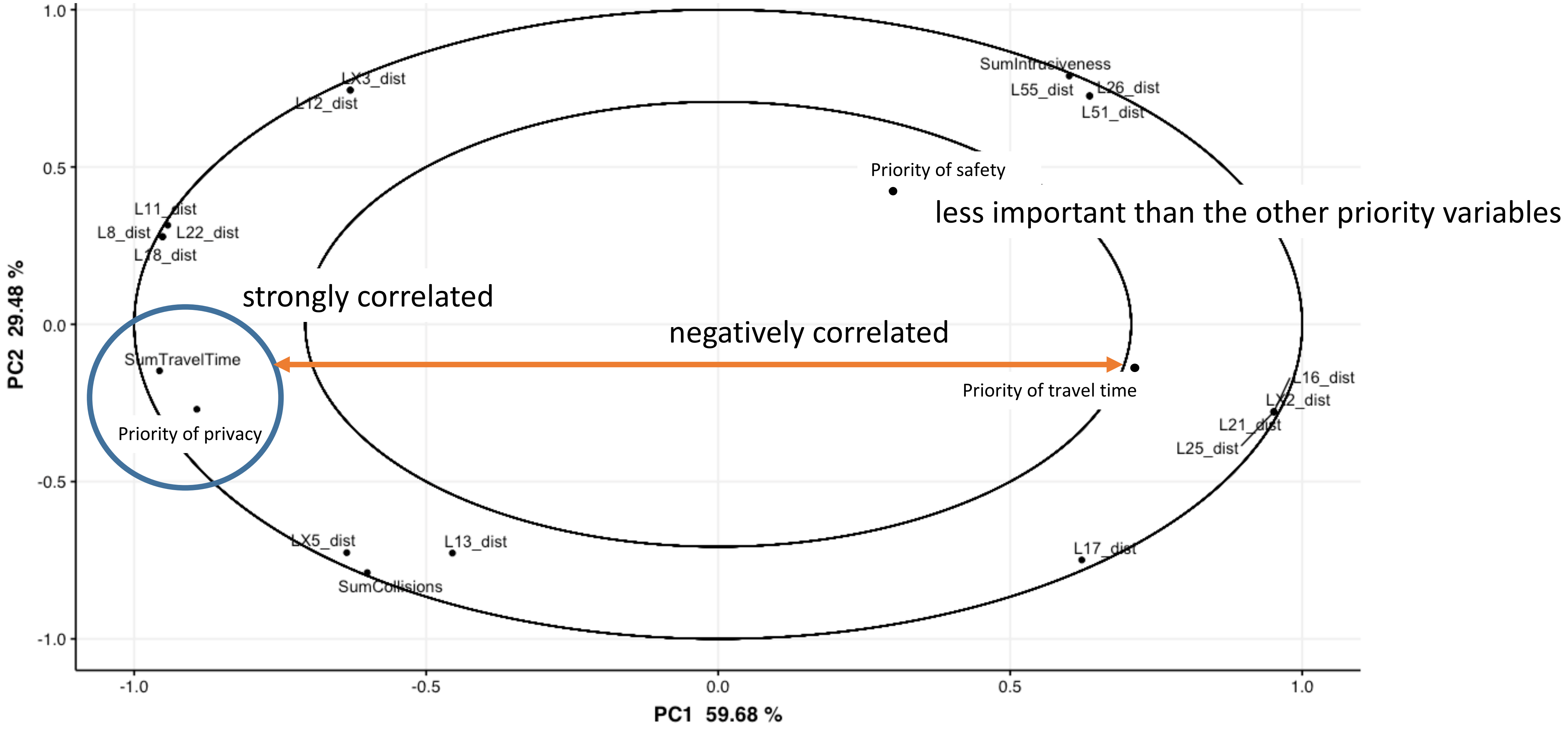
$$multi(\mathcal{R}^{c}_{\min}[C], \mathcal{P}_{\geq 1}[F⑥ \wedge G(\neg ⑤)],\ \mathcal{R}^{c_t}_{\leq 5}[C])$$

Generating data...

priority of travel time	priority of safety	priority of privacy	cost of travel time	cost of safety	cost of privacy	number of steps	decision at Location L1	decision at Location L2	decision at Location L3	...
0	0	1	50	100	10	25	L2	L4		...
0.1	0	0.9	50	100	10	25	L2	L4		...
0.2	0	0.8	50	100	10	25	L2	L4		...
0.3	0	0.7	45	96	20	22	L3		L6	...
...										

Wohlrab, R., Cámara, J., Garlan, D., & Schmerl, B. (2023). Explaining quality attribute tradeoffs in automated planning for self-adaptive systems. *Journal of Systems and Software.*

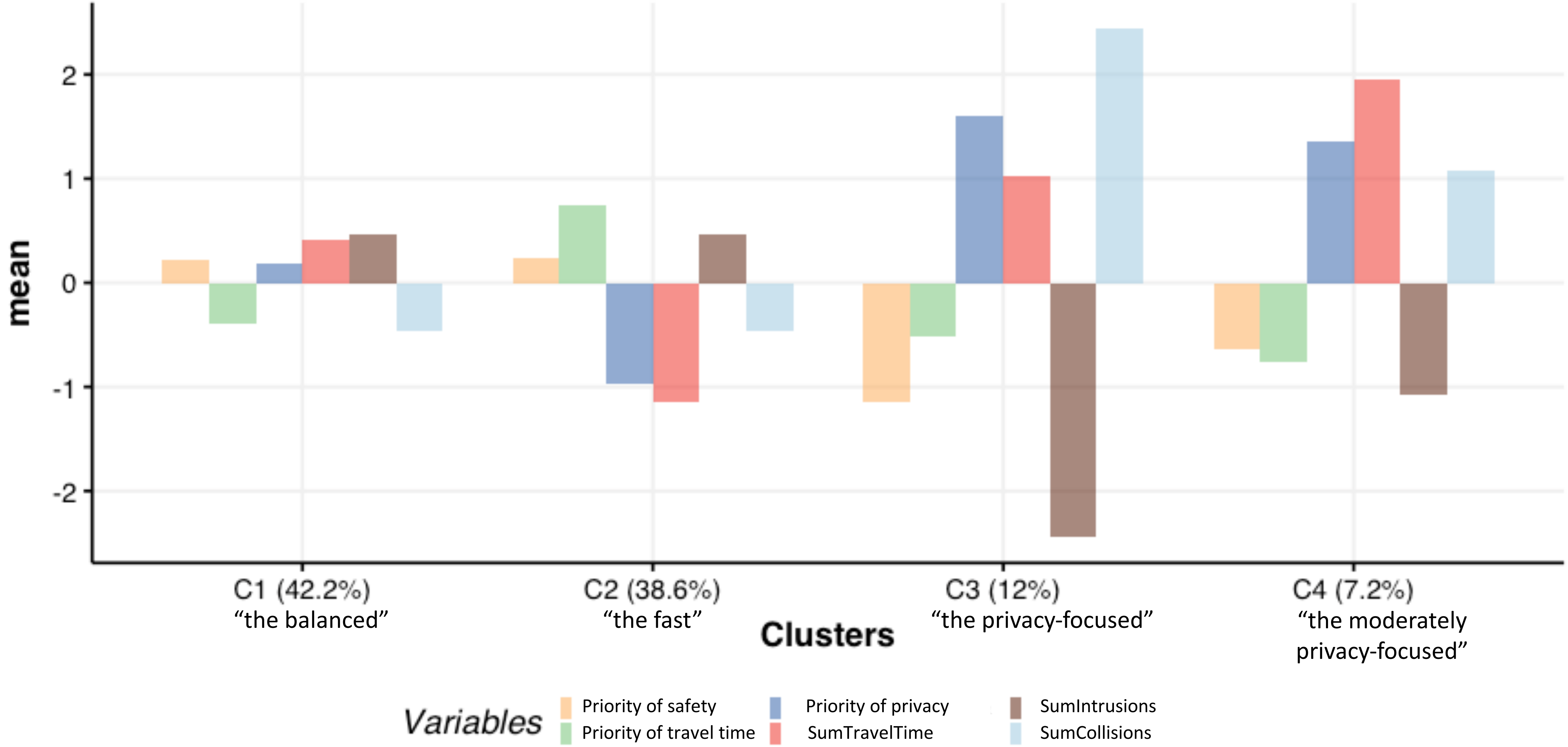
Principal Component Analysis (PCA)



Generating data...

priority of travel time	priority of safety	priority of privacy	cost of travel time	cost of safety	cost of privacy	number of steps	decision at Location L1	decision at Location L2	decision at Location L3	...
0	0	1	50	100	10	25	L2	L4		...
0.1	0	0.9	50	100	10	25	L2	L4		...
0.2	0	0.8	50	100	10	25	L2	L4		...
0.3	0	0.7	45	96	20	22	L3		L6	...
...										

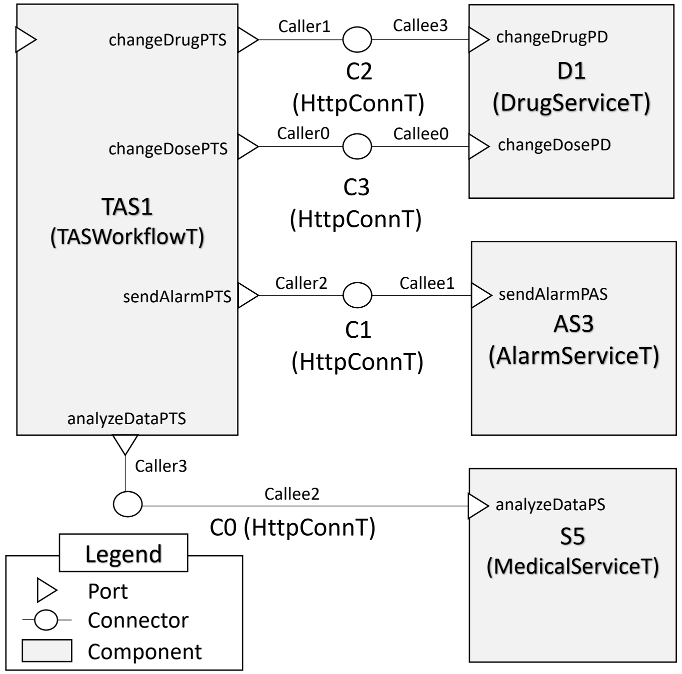
Wohlrab, R., Cámara, J., Garlan, D., & Schmerl, B. (2023). Explaining quality attribute tradeoffs in automated planning for self-adaptive systems. *Journal of Systems and Software.*



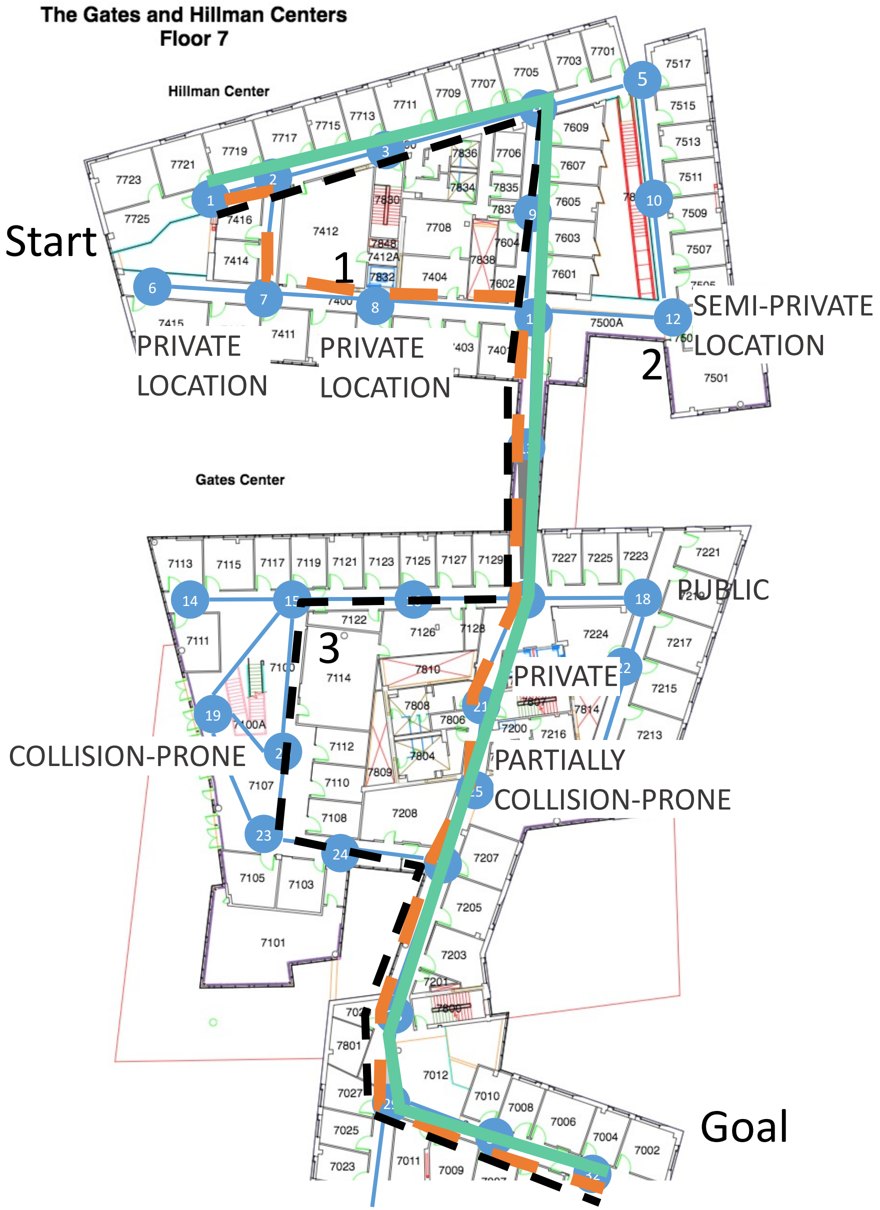
Wohlrab, R., Cámara, J., Garlan, D., & Schmerl, B. (2023). Explaining quality attribute tradeoffs in automated planning for self-adaptive systems. *Journal of Systems and Software.*

Other application domains for tradeoff explanations

- Software architecture
- Smart manufacturing
- Cloud-based systems design (thesis with Volvo Cars)

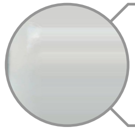


Cámara, J., Wohlrab, R., Garlan, D., & Schmerl, B. (2023). ExTrA: Explaining architectural design tradeoff spaces via dimensionality reduction. *Journal of Systems and Software*, *198*, 111578.
Cámara, J., Wohlrab, R., Garlan, D., & Schmerl, B. (2023). Focusing on What Matters: Explaining Quality Tradeoffs in Software-Intensive Systems via Dimensionality Reduction. *IEEE Software*.
Garlan, D., Schmerl, B., Wohlrab, R. & Cámara, J. (2024). *Challenges in Creating Effective Automated Design Environments: an experience report from the domain of generative manufacturing.* In *Designing*.



Qualities in automated mission planning


Safety (collision avoidance)


Privacy


Efficiency (travel time)

- Problems:
  - Autonomous systems often do not allow for changing priorities of objectives at runtime
  - Tradeoffs tend to be opaque

Option 1:
Travel time
Safety
Privacy

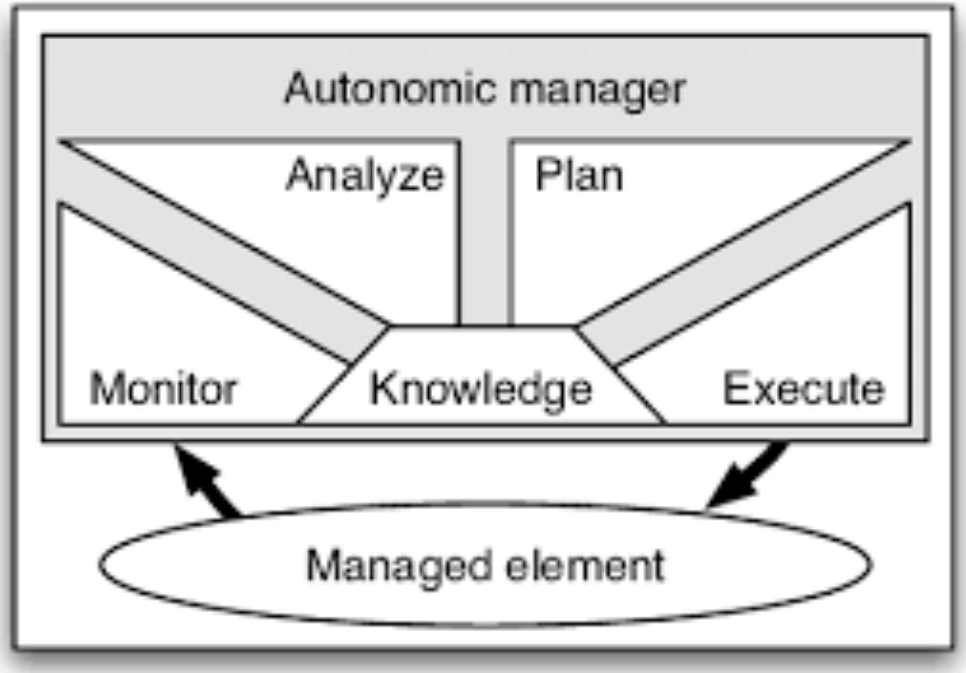
Option 2:
Travel time
Safety
Privacy

Option 3:
Travel time
Safety
Privacy

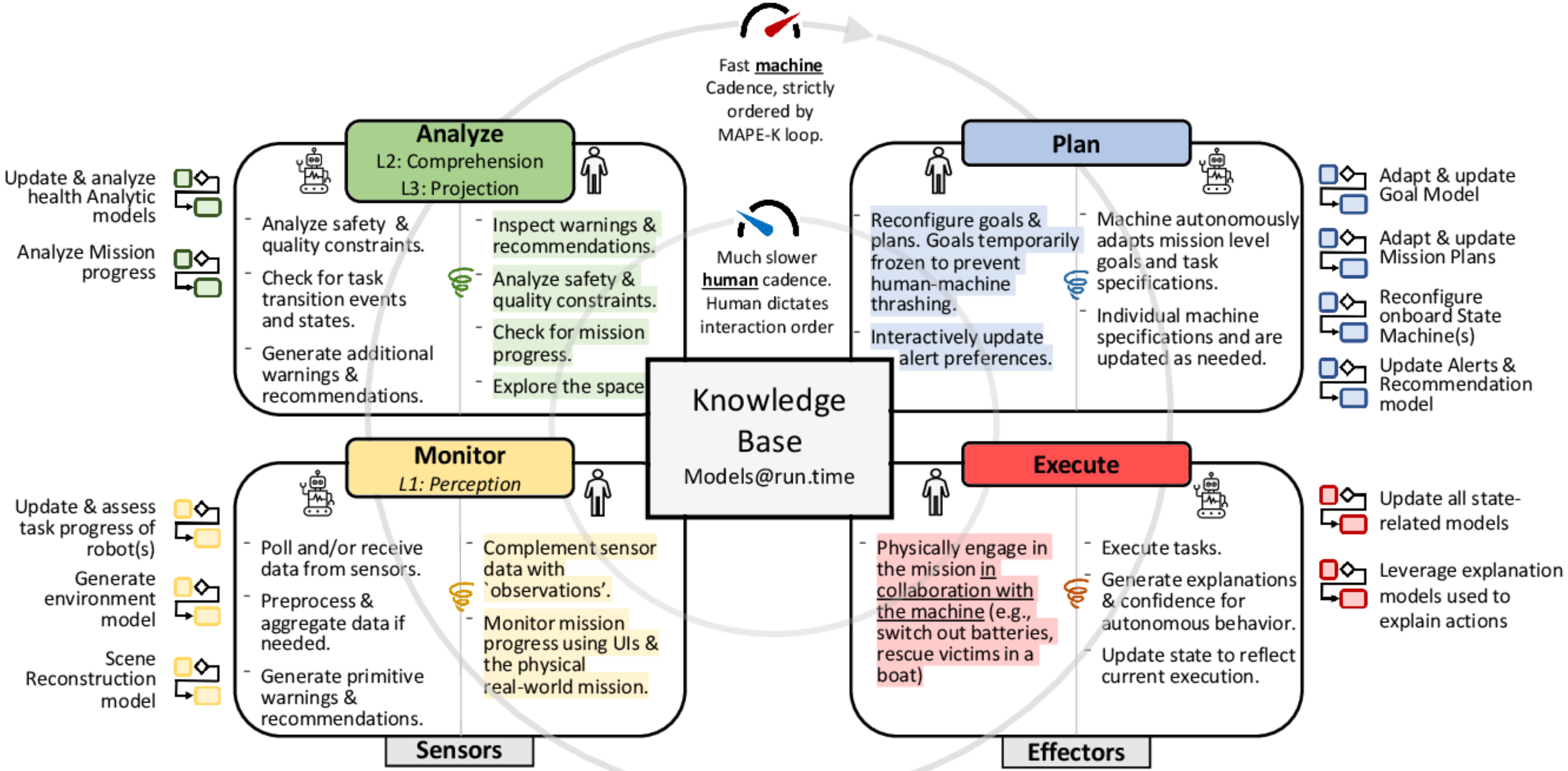
Wohlrab, R., Cámara, J., Garlan, D., & Schmerl, B. (2023). Explaining quality attribute tradeoffs in automated planning for self-adaptive systems. *Journal of Systems and Software*, *198*, 111538.

Self-adaptive systems

- A self-adaptive system is a system that can handle changes and uncertainties in its environment, the system itself, and its goals autonomously.

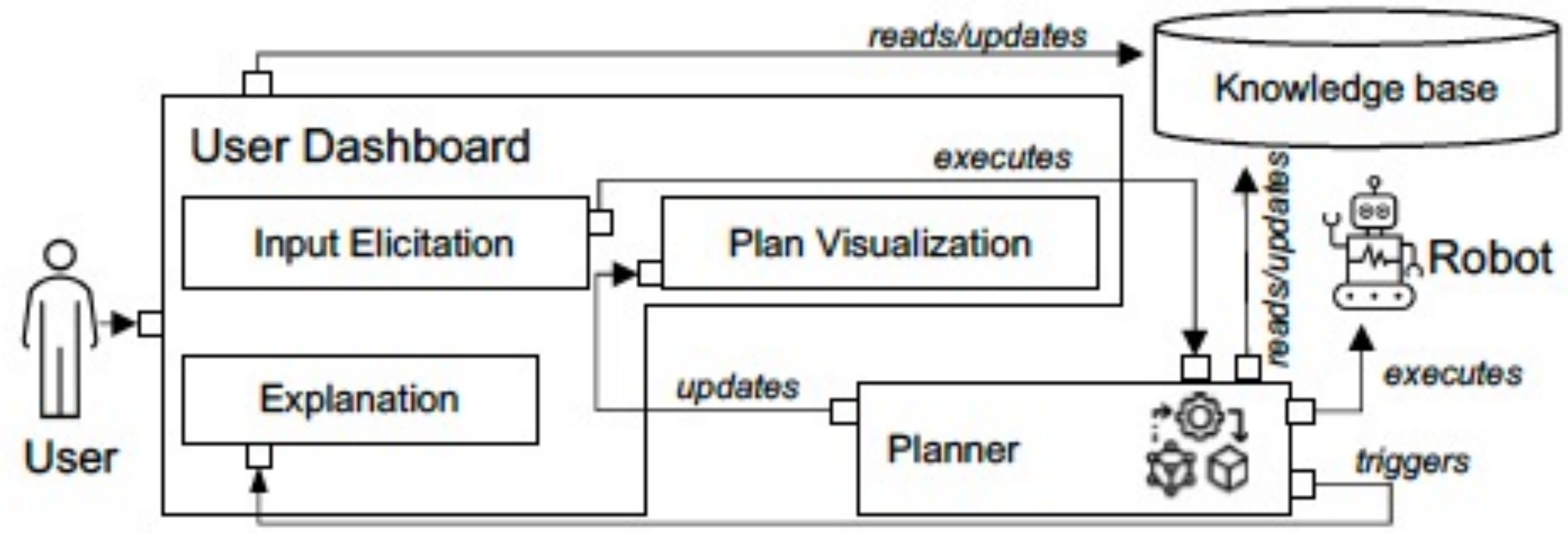


Weyns, D. (2020). *An introduction to self-adaptive systems: A contemporary software engineering perspective*. John Wiley & Sons.
Kephart, J. O., & Chess, D. M. (2003). The vision of autonomic computing. *Computer*, *36*(1), 41-50.



Cleland-Huang, J., Agrawal, A., Vierhauser, M., Murphy, M., & Prieto, M. (2022). Extending MAPE-K to support human-machine teaming. In *SEAMS* (pp. 120-131).

Architecture: Elicitation and Explanation Framework

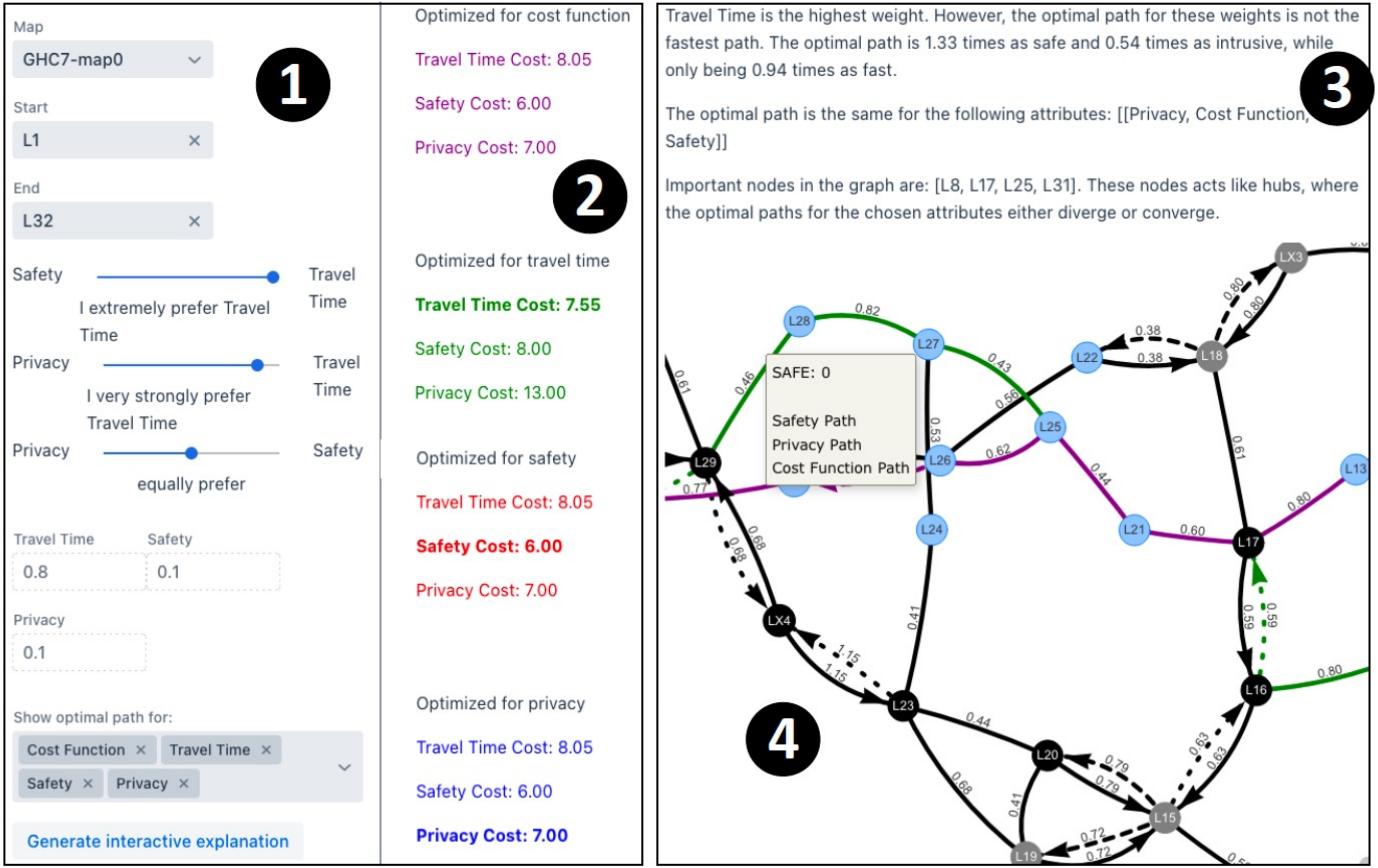


Wohlrab, R., Vierhauser, M. & Nilsson, E. (2024). What Impact do my Preferences Have? A Framework for Explanation-Based Elicitation of Quality Objectives for Robotic Mission Planning. In *30th International Working Conference on Requirement Engineering: Foundation for Software Quality.*

Participants

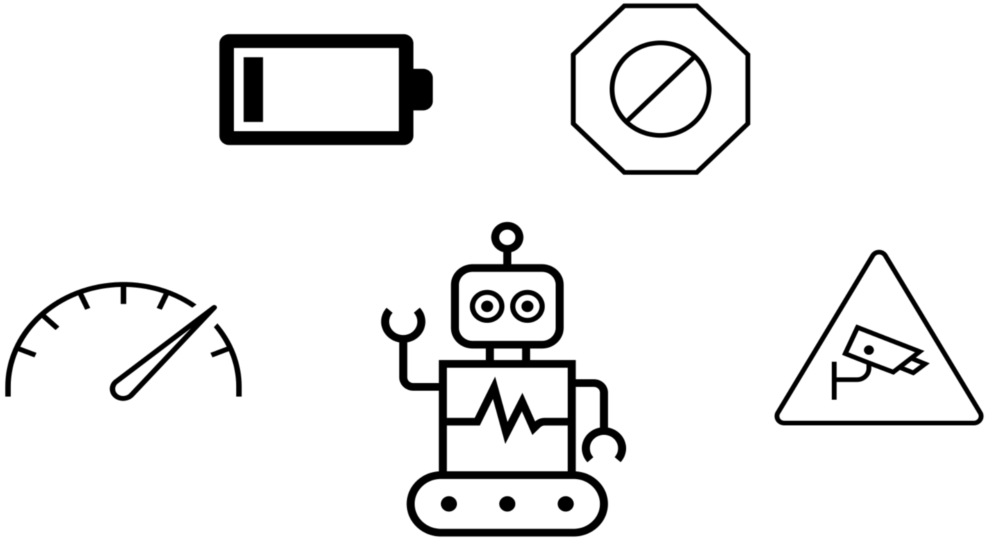
Part.	Occupation	Experience with technical jects
1	Engineering manager	3-5 yrs.
2	Graduating software development student	1-2 yrs.
3	Graduating software engineering student	1-2 yrs.
4	Software developer	6+ yrs.
5	Backend developer	3-5 yrs.
6	Cloud engineer/architect	6+ yrs.
7	UX-design student	0 yrs.
8	Consultant manager	6+ yrs.
9	Software architect	6+ yrs.
10	Product owner	6+ yrs.

Wohlrab, R., Vierhauser, M. & Nilsson, E. (2024). What Impact do my Preferences Have? A Framework for Explanation-Based Elicitation of Quality Objectives for Robotic Mission Planning. In *30th International Working Conference on Requirement Engineering: Foundation for Software Quality.*

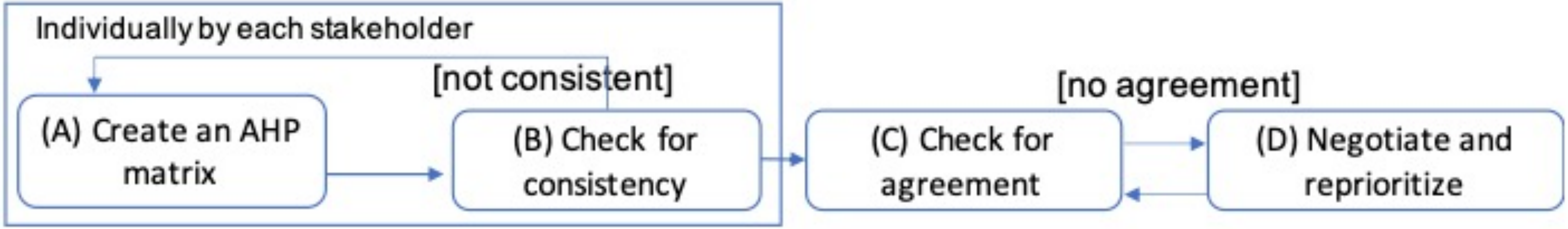

Map
GHC7-map0
Start
L1
End
L32
Safety
Travel Time
I extremely prefer Travel Time
Privacy
Travel Time
I very strongly prefer Travel Time
Privacy
Safety
equally prefer
Travel Time
0.8
Safety
0.1
Privacy
0.1
Show optimal path for:
Cost Function
Travel Time
Safety
Privacy
Generate interactive explanation
1
Optimized for cost function
Travel Time Cost: 8.05
Safety Cost: 6.00
Privacy Cost: 7.00
2
Optimized for travel time
Travel Time Cost: 7.55
Safety Cost: 8.00
Privacy Cost: 13.00
Optimized for safety
Travel Time Cost: 8.05
Safety Cost: 6.00
Privacy Cost: 7.00
Optimized for privacy
Travel Time Cost: 8.05
Safety Cost: 6.00
Privacy Cost: 7.00
Travel Time is the highest weight. However, the optimal path for these weights is not the fastest path. The optimal path is 1.33 times as safe and 0.54 times as intrusive, while only being 0.94 times as fast.
3
The optimal path is the same for the following attributes: [[Privacy, Cost Function, Safety]]
Important nodes in the graph are: [L8, L17, L25, L31]. These nodes acts like hubs, where the optimal paths for the chosen attributes either diverge or converge.
SAFE: 0
Safety Path
Privacy Path
Cost Function Path
4
LX3
L28
L27
L22
L18
L25
L26
L29
L24
L21
L17
L13
LX4
L23
L16
L20
L15
L19
0.82
0.46
0.43
0.38
0.80
0.56
0.53
0.62
0.61
0.44
0.77
0.68
0.60
0.59
0.41
1.15
0.44
0.79
0.63
0.72

Method for Utility Function Definition

- Based on the Analytic Hierarchy Process (AHP) – pairwise comparison of quality attributes



$U(m) = 0.8 \cdot safety(m) + 0.1 \cdot duration(m) + 0.1 \cdot privacy(m)$

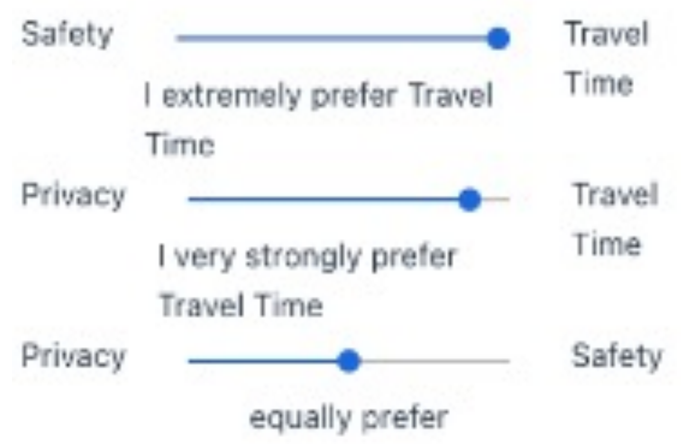


Saaty, R.: The analytic hierarchy process—what it is and how it is used. Mathematical Modelling 9(3), 161–176 (1987)
Wohlrab, R., & Garlan, D. (2023). A negotiation support system for defining utility functions for multi-stakeholder self-adaptive systems. *Requirements Engineering*, *28*(1), 3-22.

Create an AHP matrix

	Travel time	Safety	Privacy
Travel time	1	9	7
Safety	1/9	1	1
Privacy	1/7	1	1

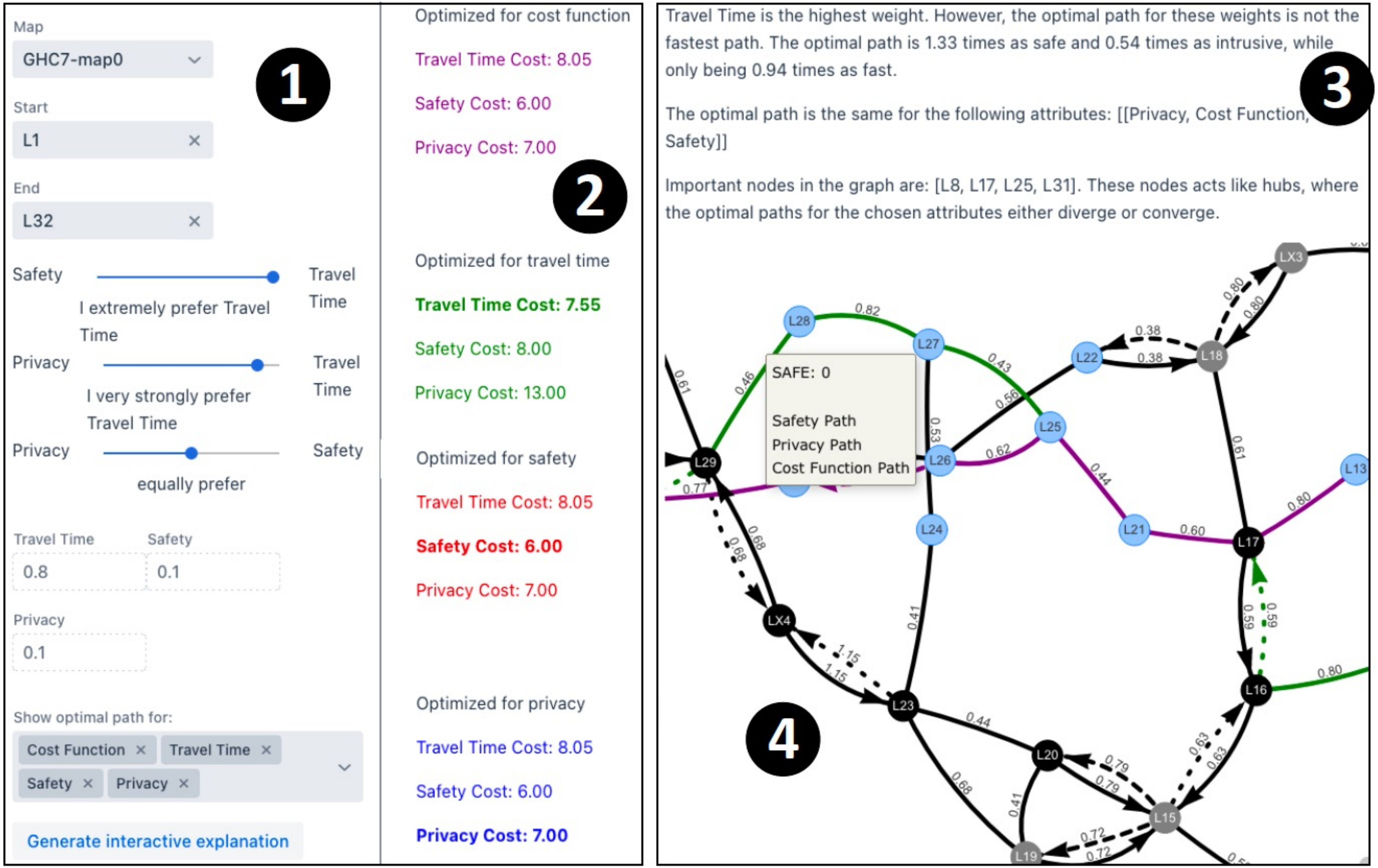
- Pairwise comparison of qualities
  - Extremely preferred 9
    Very strongly preferred 7
    Strongly preferred 5
    Moderately preferred 3
    Equally preferred 1
  - Creation of a reciprocal matrix A
- Normalized principal eigenvector of the matrix A represents the relative priorities of the qualities



Principal eigenvalue: $\lambda_{max} \approx 3.01$

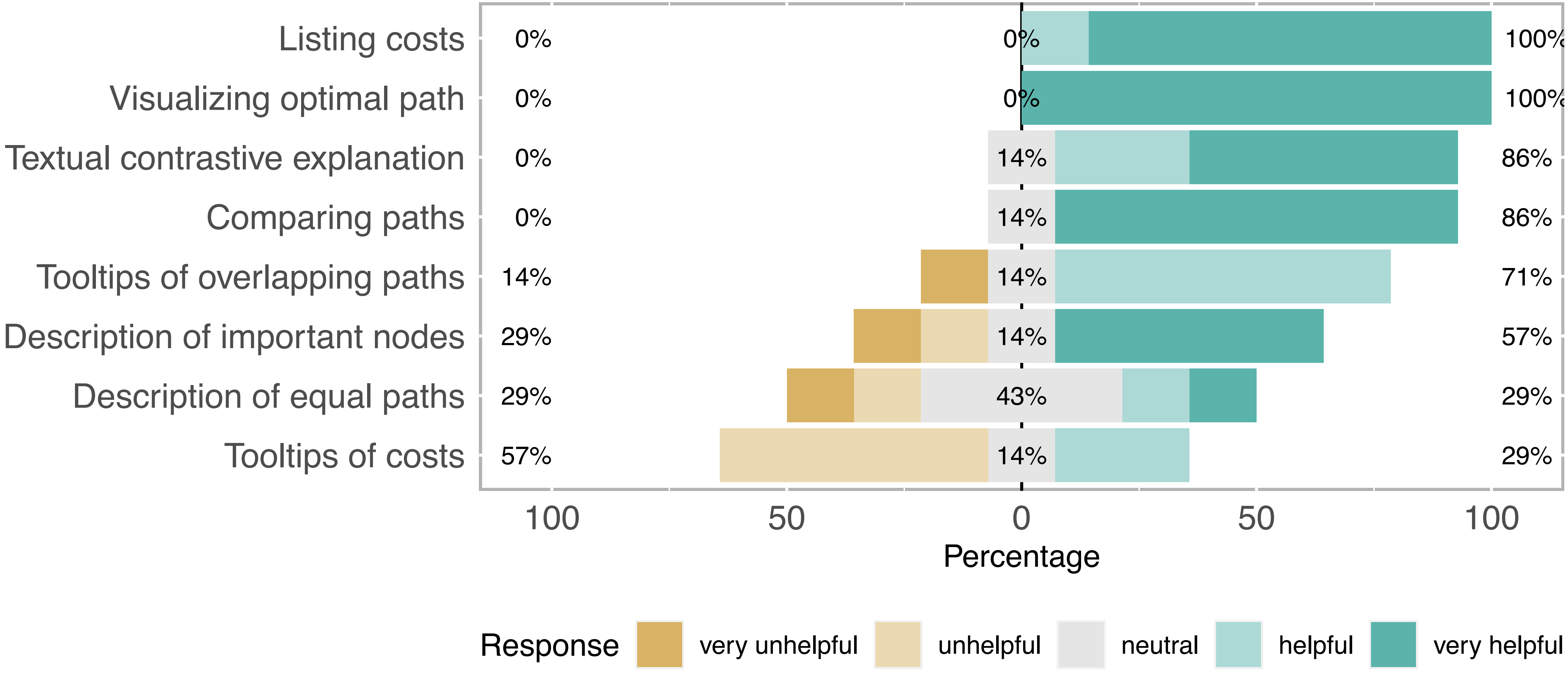
Corresponding normalized eigenvector:
(0.799, 0.105, 0.096)

U(m) = 0.799·safety(m)+0.105·duration(m)+0.096·energy(m)



1
Map
GHC7-map0
Start
L1
End
L32
Safety
Travel Time
I extremely prefer Travel Time
Privacy
Travel Time
I very strongly prefer Travel Time
Privacy
Safety
equally prefer
Travel Time
0.8
Safety
0.1
Privacy
0.1
Show optimal path for:
Cost Function
Travel Time
Safety
Privacy
Generate interactive explanation
2
Optimized for cost function
Travel Time Cost: 8.05
Safety Cost: 6.00
Privacy Cost: 7.00
Optimized for travel time
Travel Time Cost: 7.55
Safety Cost: 8.00
Privacy Cost: 13.00
Optimized for safety
Travel Time Cost: 8.05
Safety Cost: 6.00
Privacy Cost: 7.00
Optimized for privacy
Travel Time Cost: 8.05
Safety Cost: 6.00
Privacy Cost: 7.00
3
Travel Time is the highest weight. However, the optimal path for these weights is not the fastest path. The optimal path is 1.33 times as safe and 0.54 times as intrusive, while only being 0.94 times as fast.
The optimal path is the same for the following attributes: [[Privacy, Cost Function, Safety]]
Important nodes in the graph are: [L8, L17, L25, L31]. These nodes acts like hubs, where the optimal paths for the chosen attributes either diverge or converge.
4
SAFE: 0
Safety Path
Privacy Path
Cost Function Path

Evaluation



Listing costs
Visualizing optimal path
Textual contrastive explanation
Comparing paths
Tooltips of overlapping paths
Description of important nodes
Description of equal paths
Tooltips of costs
0%
0%
100%
0%
0%
100%
0%
14%
86%
0%
14%
86%
14%
14%
71%
29%
14%
57%
29%
43%
29%
57%
14%
29%
100
50
0
50
100
Percentage
Response
very unhelpful
unhelpful
neutral
helpful
very helpful

What's next?

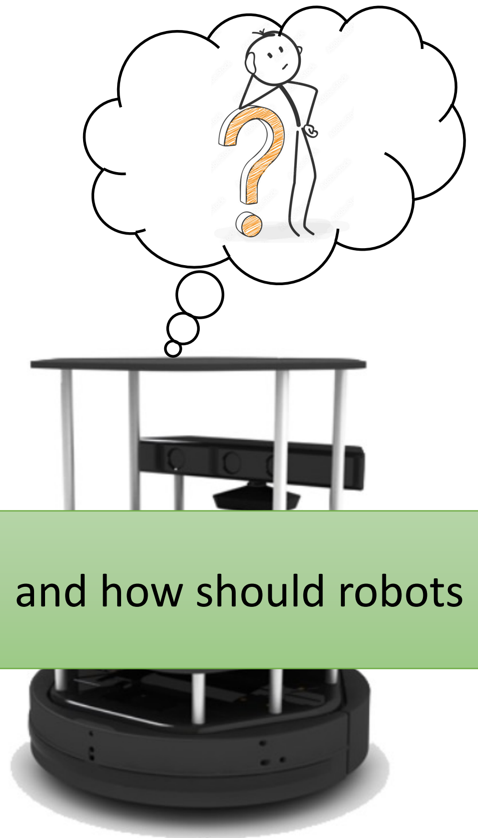
1) When should systems provide what kinds of explanations?
(in runtime robot mission planning, when cybersecurity attacks happen, when developers make design decisions, ...)

Monitoring [4]

Explainability [1]

What does the human do?

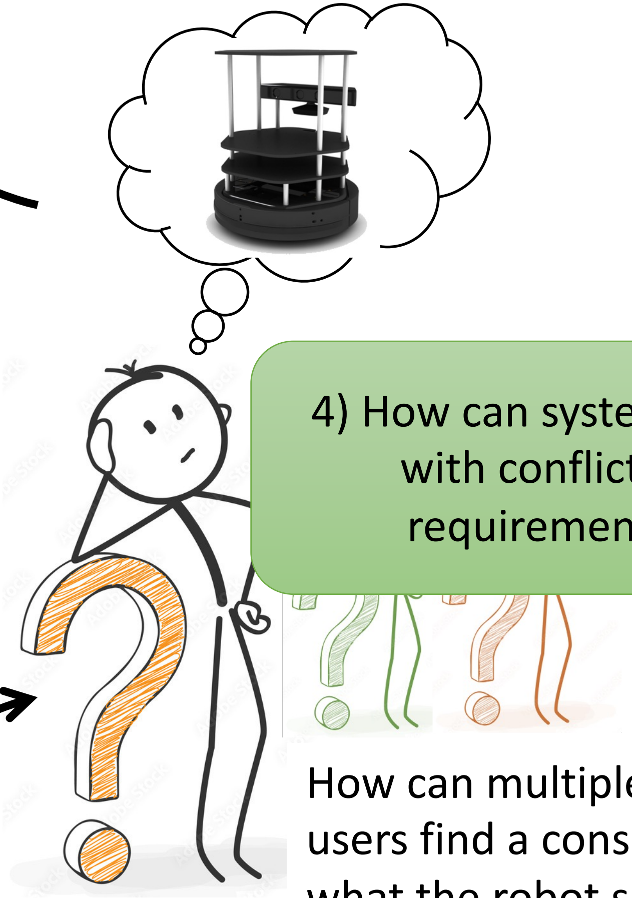
What does the robot do?



What does the human want the robot to do?

Preference elicitation [2, 3]

2) How can systems elicit humans' preferences in different contexts?



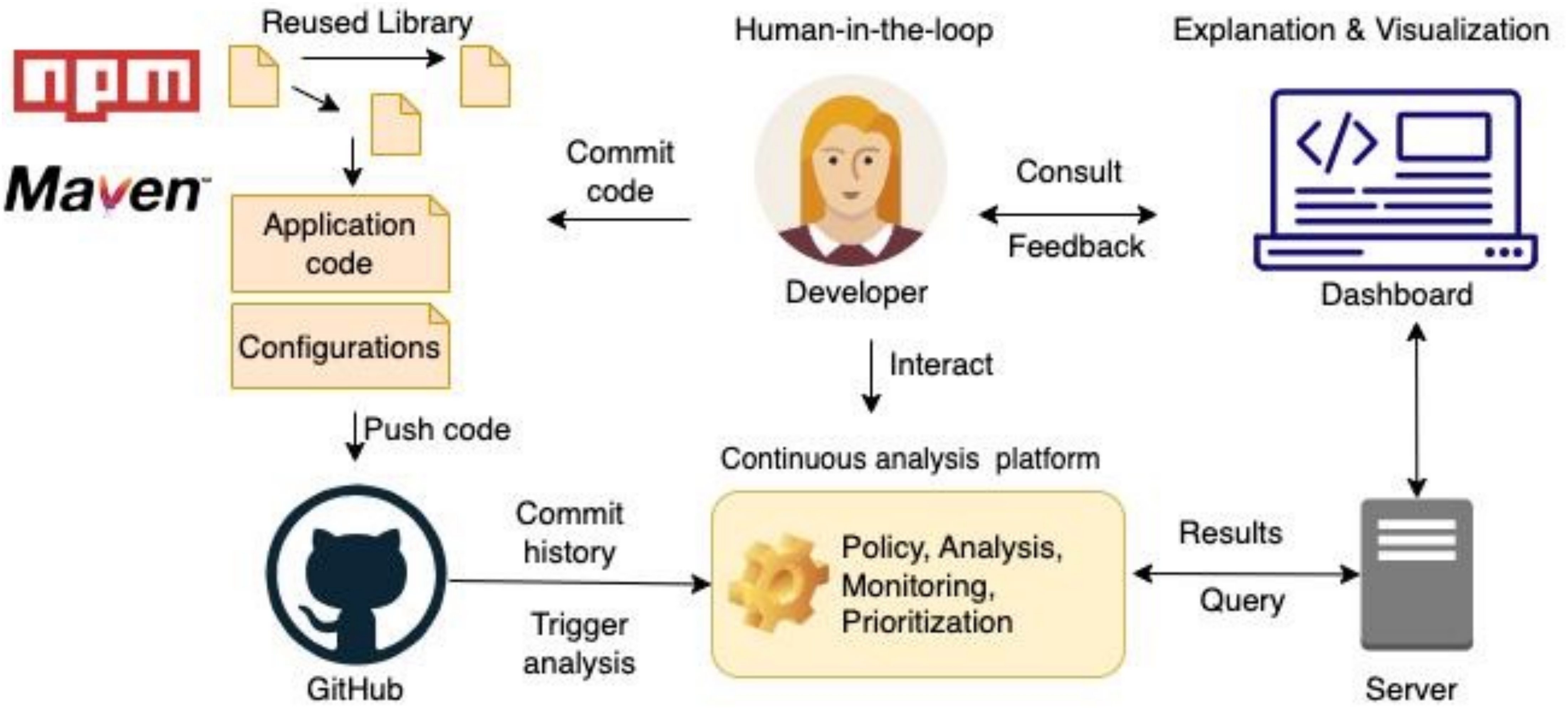
4) How can systems deal with conflicting requirements?

3) When and how should robots initiate collaborative tasks?

What does the robot want the human to do?

How can multiple human users find a consensus on what the robot should do?

ShiftLeft – a WASP NEST project



Thank you!

Rebekka Wohlrab
wohlrab@chalmers.se