

Interactive Explanations of Agent Behavior

Yotam Amitai

Technion
yotama@campus.technion.ac.il



Guy Avni

University of Haifa
gavni@cs.haifa.ac.il

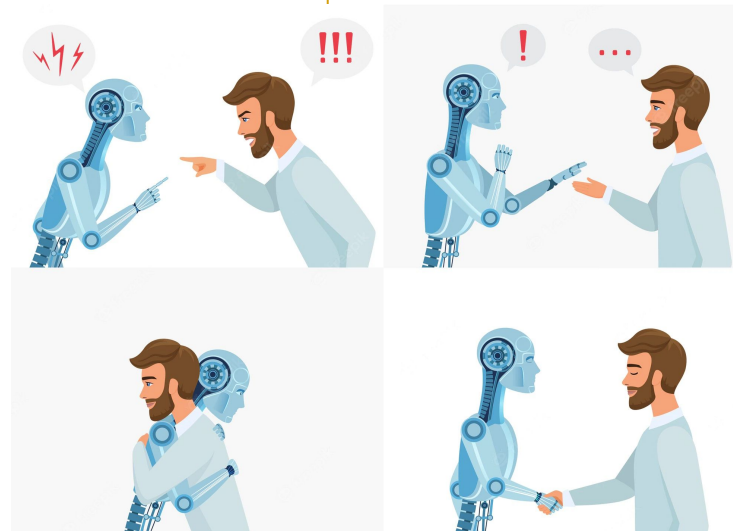
Ofra Amir

Technion
oamir@technion.ac.il



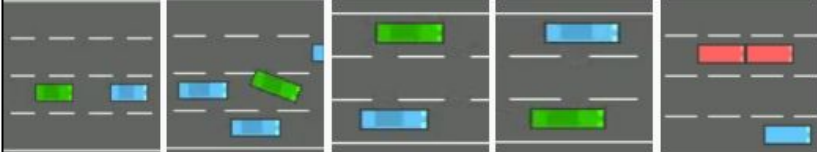
Interactive Explanations of Agent Behavior

- XAI is crucial
- Previous approaches are static
- Advocate knowledge to be **sought**
- Explanation by demonstration



Agent System Queries - Interactive Tool (ASQ-IT)

Behind: In front of: Above: Below: Collision:



Start Frame:

Position Lane


End Frame:

Position Lane

Constraints:

Constraint

Element A Element B

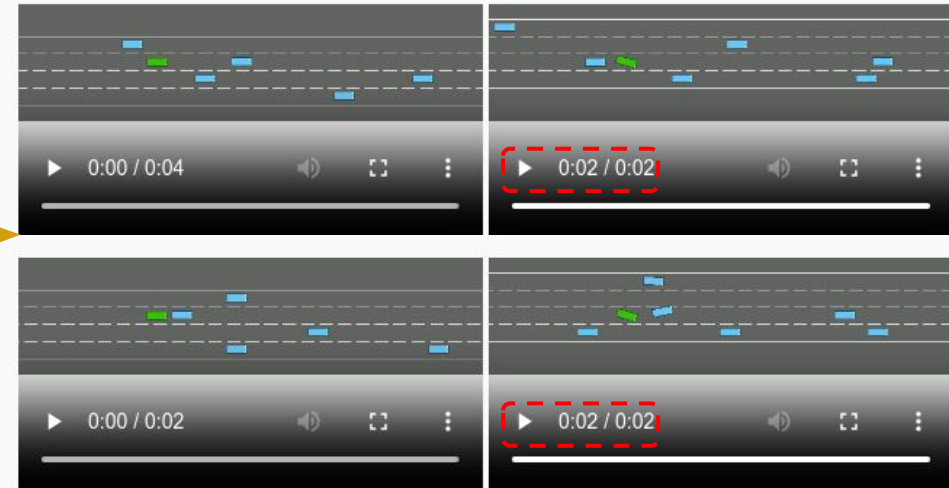


Your videos have been generated! 4 more videos are available.

Your Query:

Start Frame: behind lane2, End Frame: lane3

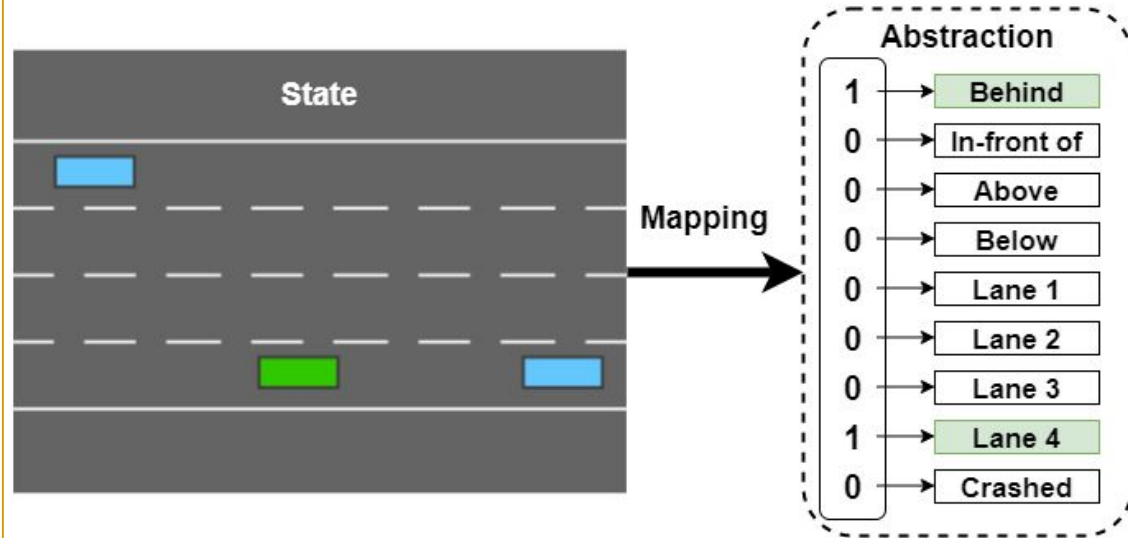
Videos matching your query:



[Load More Videos](#)

Abstracting The State

- Discretizing key elements
 - Iterative process
- Save trajectories to create interaction library

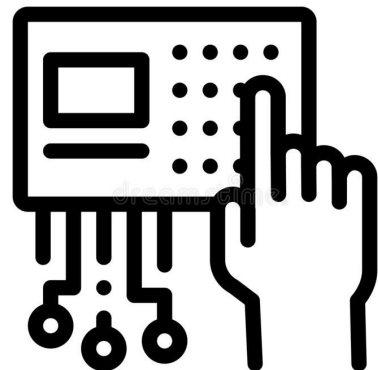


The Algorithm

Formal problem: Given a trace and an LTLf formula, find (and show) substraces that satisfy the formula.

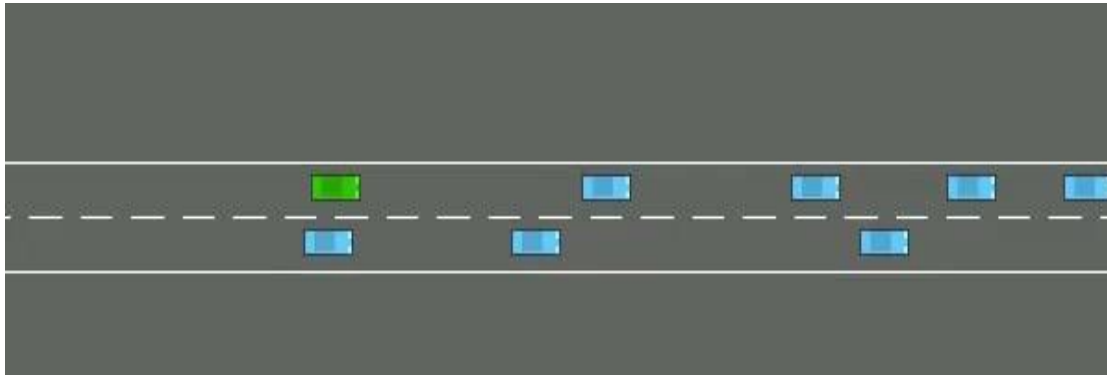
Given a query (from UI):

1. Translate query to LTLf formula φ
2. Construct two DFAs $A_\varphi, A_{F\varphi}$ using LTLf2DFA
3. Feed trace to $A_{F\varphi}$ until it accepts.
4. Feed backwards to A_φ to find suffix beginning that satisfies φ
5. Restart from last reached index.



User Study 1: Usability Assessment

- Examine laypeople interaction with ASQ-IT
 - 40 participants via Prolific, payment - 4.5\$



- Key results:
 - Semantics Comprehension
 - Meaningful query formulation
 - Fast learning curve

Select possible specification: (Only 1 is valid)

Start Frame = Lane 1 & Above

End Frame = Lane 2

Constraint = In front of stays constant

A

Start Frame = Above

End Frame = Below & Lane 2

Constraint = Lane 2 Changes

B

Start Frame = Above

End Frame = Below

Constraint = Lane 1 Changes into Lane 2

C

Start Frame = Lane 1

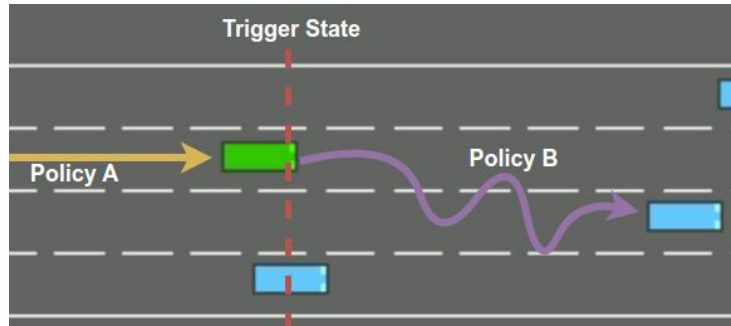
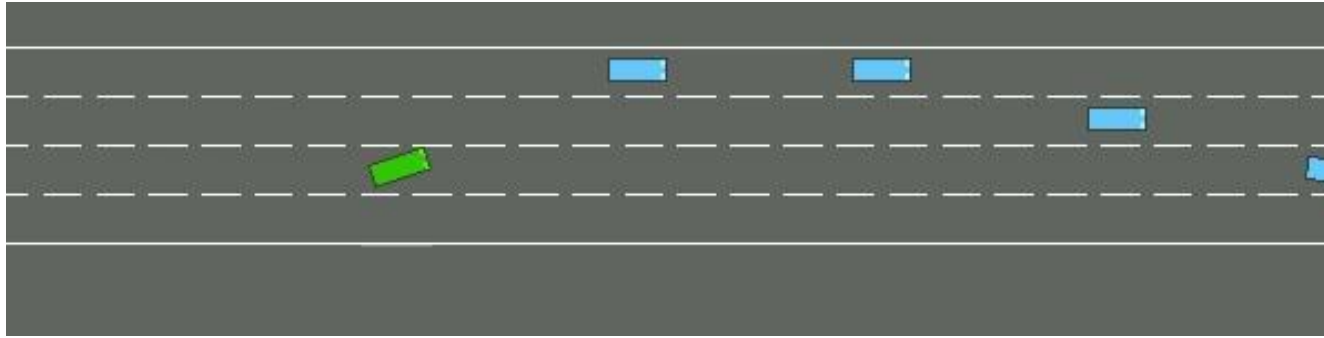
End Frame = Lane 2 & Behind

Constraint = Lane 2 Stays constant

D

User Study 2: Identifying Agent Faults

- Understand user query process and ASQ-IT usefulness
 - 13 graduate students, hour-long think-aloud interview, payment 13\$



- Key results:
 - Able to explore rigorously
 - More likely to revise & improve hypothesis
 - Higher engagement & satisfaction

Thank You

Questions?