

NEW FRONTIERS INITIATIVE SUMMER FACULTY

# LEONA: LOGIC-BASED CONTEXT-AWARE ACTIVITY INTERPRETER FOR GEOSPATIAL INTELLIGENCE

MELKIOR ORNIK

[mornik@illinois.edu](mailto:mornik@illinois.edu)



UNIVERSITY OF  
**ILLINOIS**  
URBANA-CHAMPAIGN

# FUTURE WORLD OF INTELLIGENCE

- Where a **computer** can positively identify an entity, among thousands of similar entities, just from vague, occasional information about their activities...
- Where a **computer** can determine all entities that exhibit behavior indicated by an analyst...
- Where a **computer** can identify **important** patterns of someone's activities...
- Where a **computer** can recognize large events just from tracking behaviors of individual humans...



# GRAND CHALLENGE OF (GEO)INT

**Rapid, relevant, understandable analysis**

Provides answers while the decision-makers can still act

Provides answers that are important to decision-makers

Provides answers in a way that decision-makers can understand

**Humans have the knowledge, but just don't have the time**

*We need to allow the humans to spend their time on things computers can't do*

# HUMAN EXPERTISE ↔ MACHINE COMPUTATION

Fundamental prerequisite: a common language

- **Easy** for a human to understand
- **Formal** enough for a computer to operate on
- **Expressive** enough to describe spatiotemporal features

**(Linear) temporal logic**

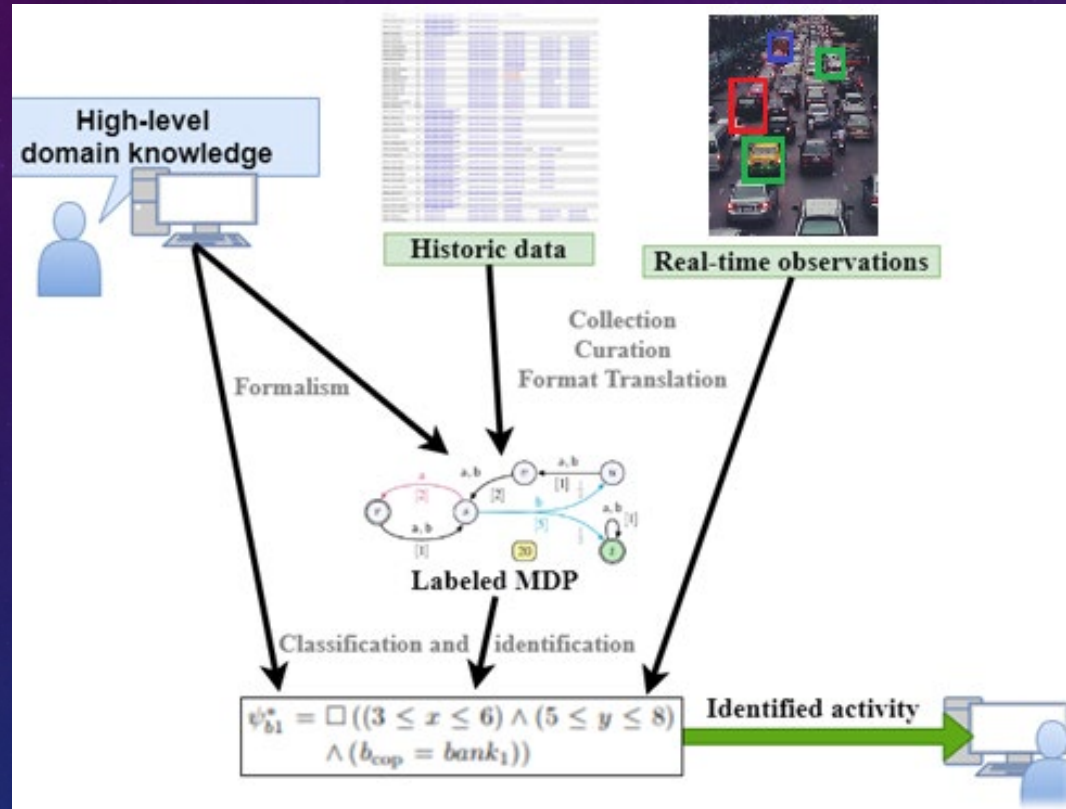
“Never go inside area A and go to area B before area C”

$$\square \neg(\underline{\text{inside A}}) \wedge \neg(\underline{\text{inside B}}) \mathcal{U}(\underline{\text{inside C}})$$

“Before the building is completed, construction work always took place only on Sundays or holidays”

$$(\underline{\text{work}} \Rightarrow (\underline{\text{Sunday}} \vee \underline{\text{holiday}})) \mathcal{U}(\underline{\text{complete}})$$

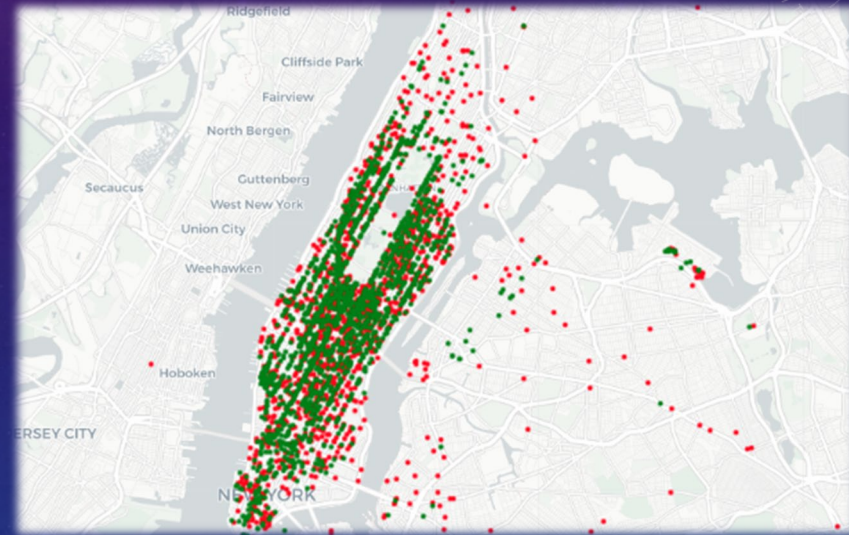
# AUTOMATED ACTIVITY IDENTIFICATION



# PROOF OF CONCEPT

All taxi trip records for New York City in 2013

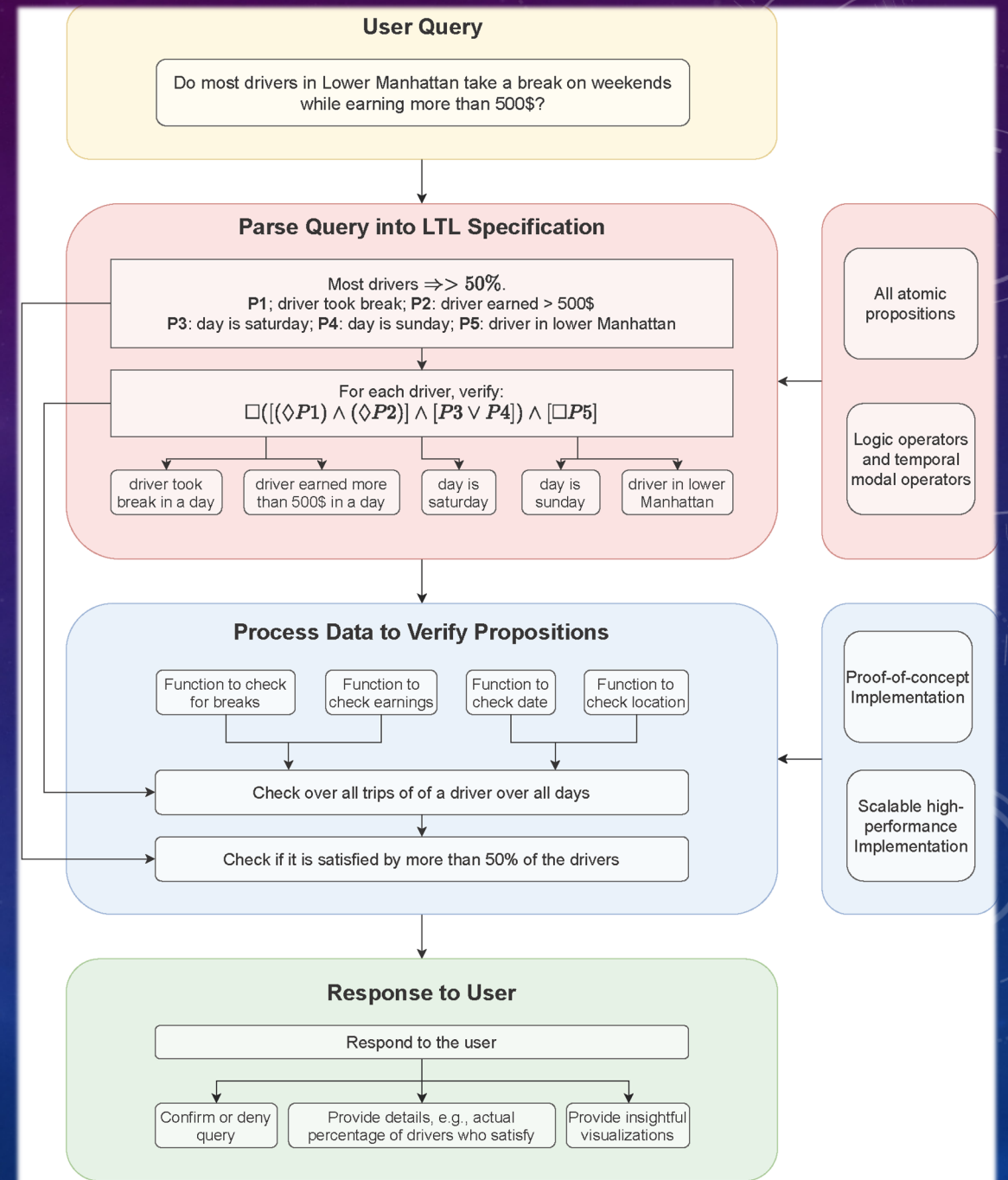
- **24 fields (28.8 GB)**, including but not limited to:
  - pick-up and drop-off times
  - pick-up and drop-off location
  - trip distance
  - itemized fare
  - payment type
  - driver-reported passenger counts
  - **pseudonymized medallion number (uniquely identifies a vehicle)**
  - **pseudonymized hack license number (uniquely identifies a driver)**



# SPECIFICATION VERIFICATION

Given a specification, determine whether it is true

- Not just “filtering” because of spatiotemporal elements
- In fully generality, boils down to formal verification on an automaton, but if the specification is easier, some of the work may be done ad hoc



# INDIVIDUAL IDENTIFICATION

Posted by u/NYCHack 9 years ago

I am a 26 year old NYC yellow cab driver. Ask me anything

In 2013, a NYC cab driver **anonymously** participated in AskMeAnything on Reddit

In the answers to users' questions, they provided some vague information about themselves:

- *I'm working full time now until the end of the summer when I'm quitting for good.*
- *Farthest place was Delaware or new haven Connecticut*
- *I picked up a guy in TriBeCa which is pretty far downtown. At the bottom of the island. The destination was all the way to Dyckman street*
- *I go to sleep around 5 AM, after my shift*

Their statements **apply to hundreds of taxi drivers**, especially when accounting for errors or imprecisions, and may **contain spatiotemporal specifications** which make manual filtering difficult



# UNIQUE IDENTIFICATION

All of the driver's relevant statements can be parsed into temporal logic

We use our pipeline:

- **relevant specifications gathered** from the Reddit post
- **context-aware “slack”** to account for imprecisions or minor errors in the statements
- **dataset** on taxi drivers obtained online
- **automated specification verification**

**The driver who participated in this AMA had (pseudonymized) hack license number  
19E2C [REDACTED] 986A4 \***

\* There is a well-known way to de-anonymize the hack license numbers. If we wanted to, we can obtain the driver's full name, even though they only gave vague statements that were true for hundreds of other drivers

# CHALLENGES

## Current capability

**Determine whether an LTL formula is correct / find a sequence of observations that satisfies it**

- Already useful for verification of **analyst “hunches”**
- Does not provide information “on its own”

## Future capability

**Determine an LTL formula that describes observed behaviors**

- Only needs an analyst notion of importance
- Possibilities of partial observability, imperfect observations
- Theoretically interesting problem **with enormous applications**

# FUTURE

We are on the brink of **automated, smart GEOINT activity identification**

**Analyst context awareness + computer “educated manual work”**

Neither can, or should, operate without the other

**mornik@illinois.edu**