

Ava: From Data to Insights Through Conversation

Rogers Jeffrey Leo John, Navneet Potti, and Jignesh M. Patel
University of Wisconsin – Madison



Rogers



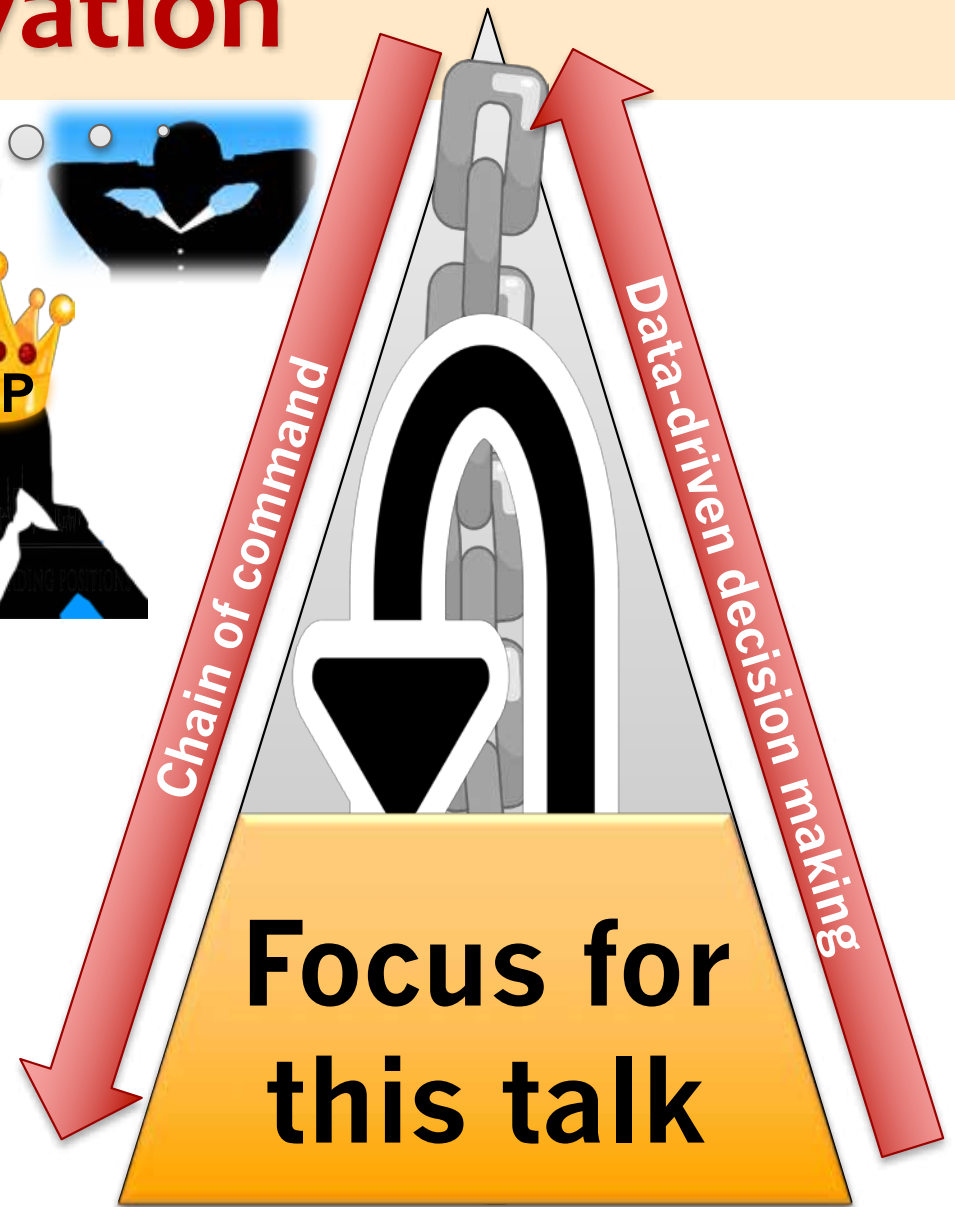
Navneet

Motivation

Why do my customers churn?



Data Scientist



A legal pyramid scheme (org chart)

Issues

1. Lost in translation
 2. Long turn-around time
 3. Correctness!
 4. Reproducibility
- ...



Data Science Pipeline



Data Loading

- from a csv file



Data Cleaning

- fill missing values



Feature Engineering

- pick/create appropriate features



Model Selection and Training

- pick an ML model based on the input and task at hand



Parameter Tuning

- hyperparameter optimization



Save model for deployment

- As a UDF, PMML file, ...

Insights

- Often the task is ...

Conversation

**Constrained
Natural
Language**

*Rely on Natural
Language
Translation rather
than Natural
Language
Understanding*

Composition

Storyboard

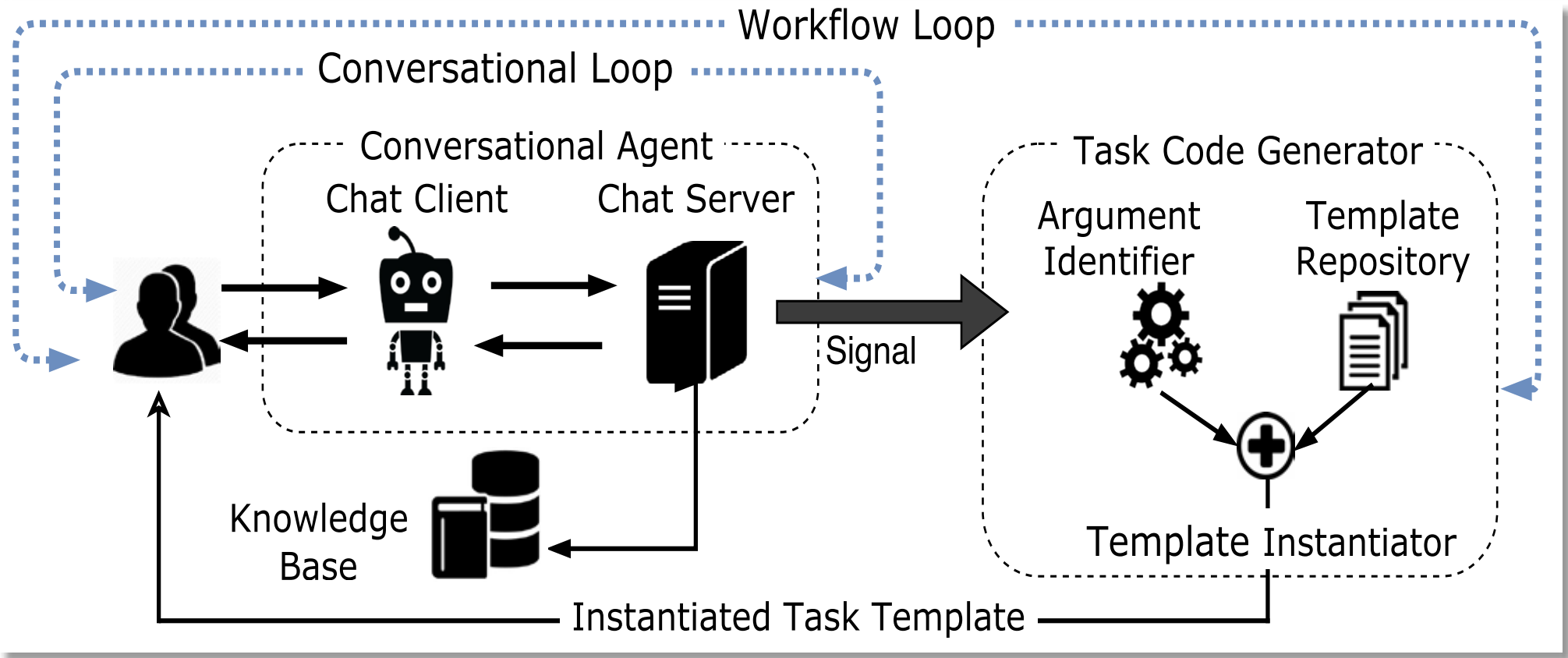
*Composable
internal architecture
and convenient
system abstraction*

Code

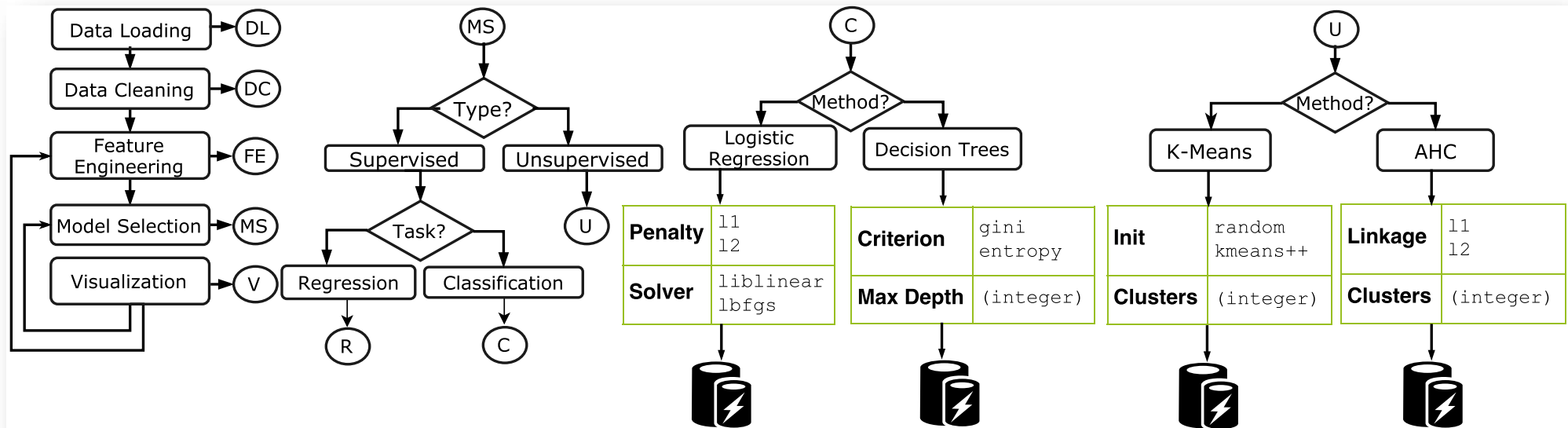
**In an interactive
notebook
(e.g. iPython)**

*Target the lower
level of the pyramid
and allow the
programmer to take
control*

Architecture



The Ava Storyboard Concept ...



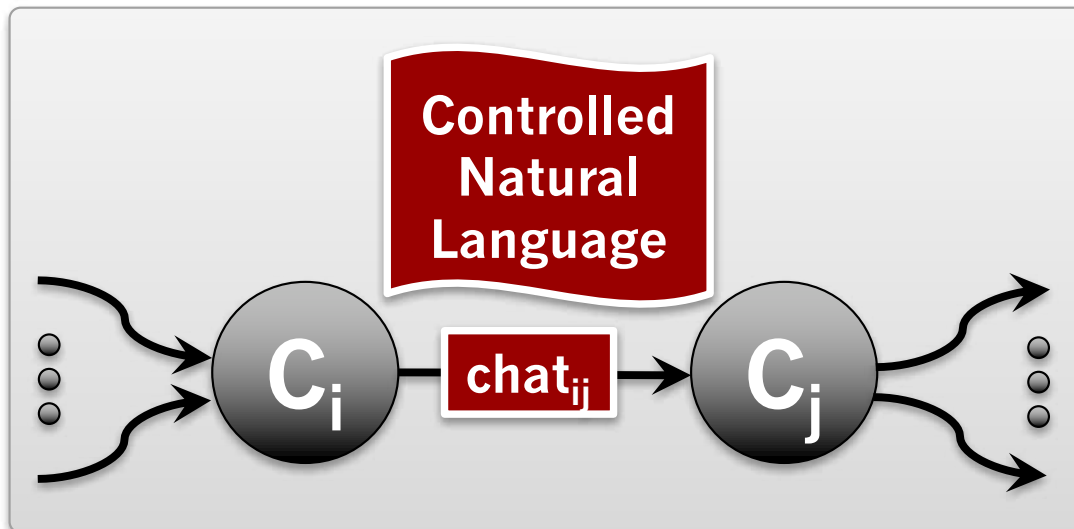
What is new?

Previous work

- Imitation Game: Turing 1950
- CNL → remove ambiguity
- Natural Language → Query:
Use feedback to refine ambiguity

Ava

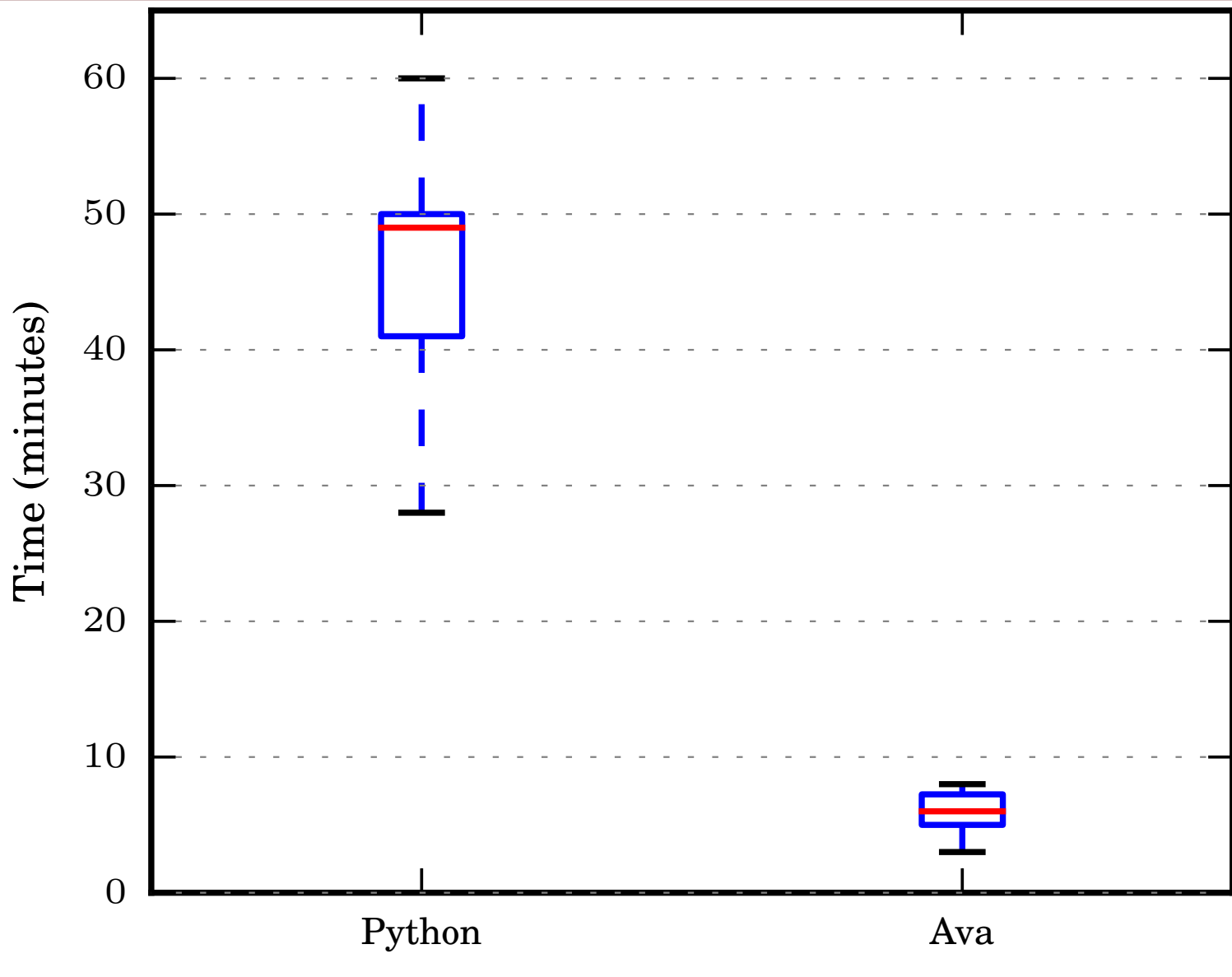
- Constrain to walk along a pre-scripted storyboard
- Storyboard is a Finite State Machine
- The user creates their “own” story. It always ends well 😊.



The right encapsulation for:

1. Technology trend
 - NLT v/s NLU
 - Different ML backends
2. Developer skills
 - Business analysts vs statistician

Results from a user study with 16 participants. Distribution of the time taken by participants to complete the first model.



Summary of AVA

The conversation is the code



The data chatbots are coming
Ava democratizes “data science”



Benefits: increased human productivity,
reproducibility, rapid exploration, a
powerful collaboration mechanism, ...



More automation along every dimension

Thanks!



Jean-Michel Ané, Victor Cabrera, Mark Craven, Pamela Herd,
David Page, Federico E. Rey, Garret Suen, and Kent Weigel

University of Wisconsin