Imperial College London

Symbolic Machine Learning for Interpretable AI

Recent advancements and future directions

Alessandra Russo

Imperial College London

Human-Like Learning

“Human thought can be seen as a model-building activity.
Human-like learning is the process of such model-building”[1]

Humans are capable of performing cognitive activities:

- Learning from past experience
  (e.g., building models of the world that can explain the observations).
- Making predictions and generating explanations
  (e.g., use learned models to generate understanding and explanation of observations)
- Revising and extending learned knowledge, based on new information
  (e.g., enable compositionality of learned models)
- Communicate their learned knowledge to others
  (e.g., support interpretability of learned models)

To realise human-like levels of cognition, Machine Learning solutions have to realise the above activities.

[1] Building machines that learn and thing like people, Brendan M. Lake, Tomer D. Ulman, Joshua B. Tenenbaum, et al., Behavioral and Brain Sciences, 40, 2017.

Imperial College London

Machine Learning

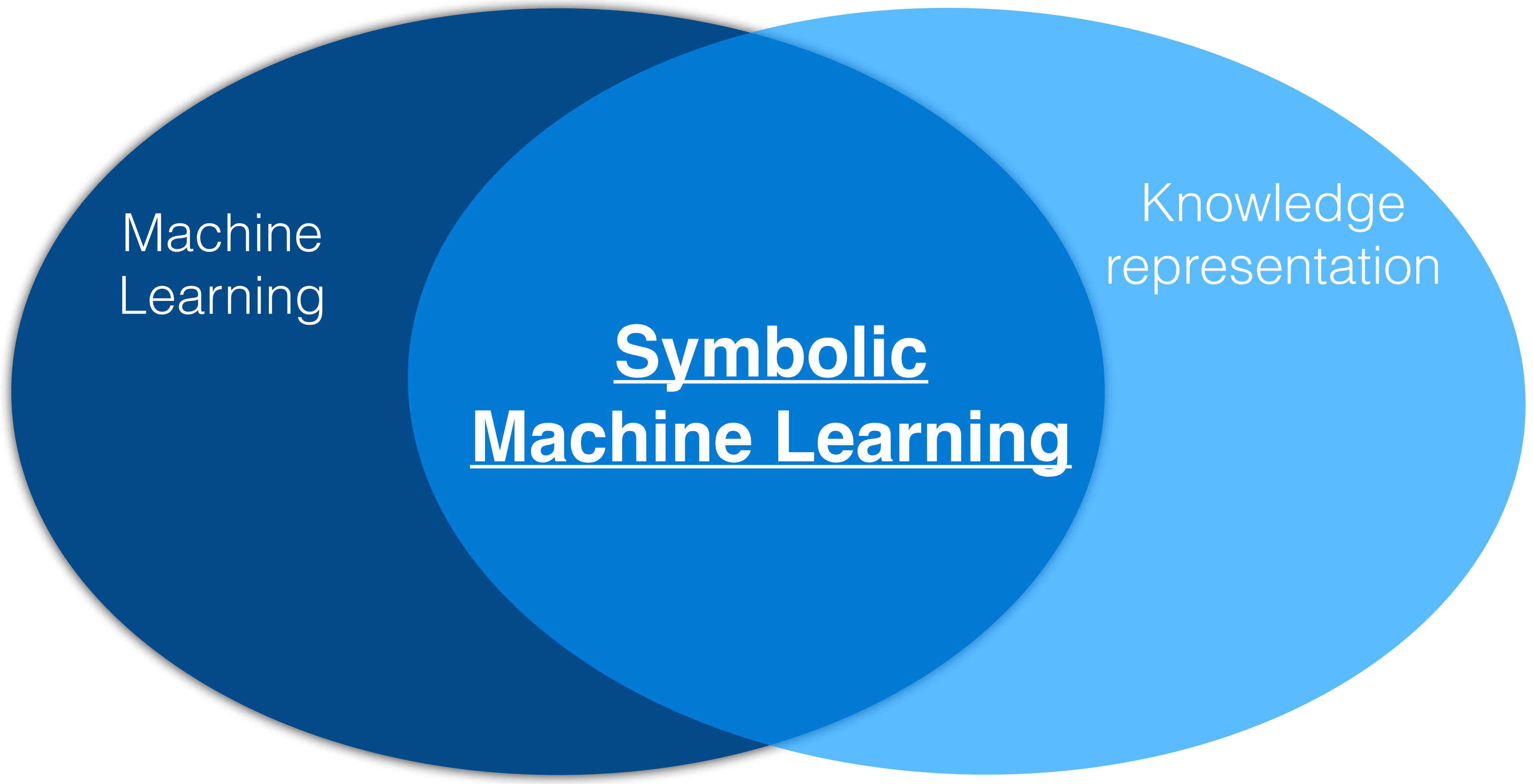
Advantages

- Ability to learn from unstructured data.
- Very effective in solving specific tasks, sometimes better than humans.

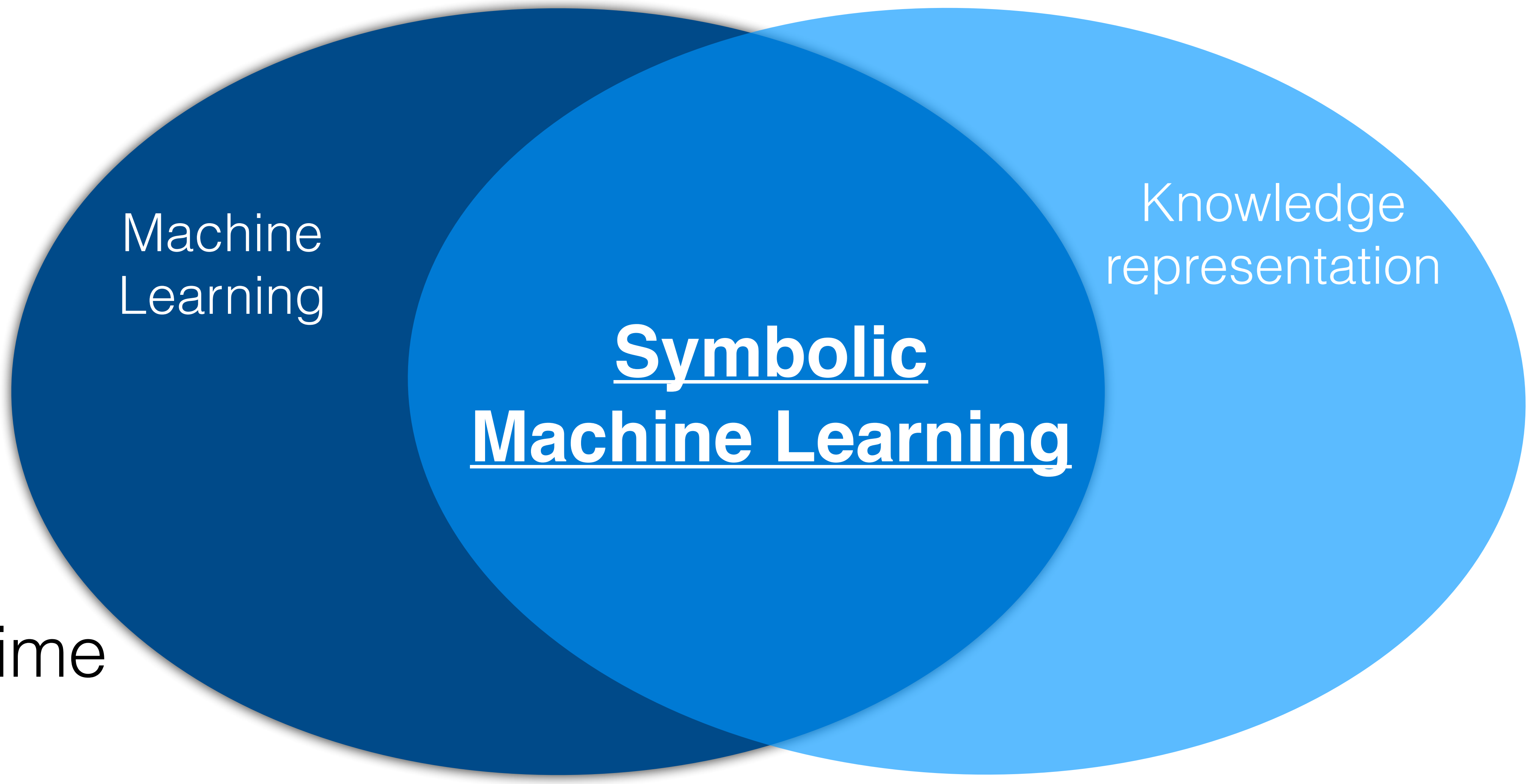
Drawbacks

- Data-intensive
- Inability to generalise
- Vulnerability to distributional shifts between training and test data
- Learned models are not interpretable
- Cannot use prior (or learned) knowledge

Symbolic Machine Learning



Symbolic Machine Learning



Information extraction from data

Predictions about unseen data

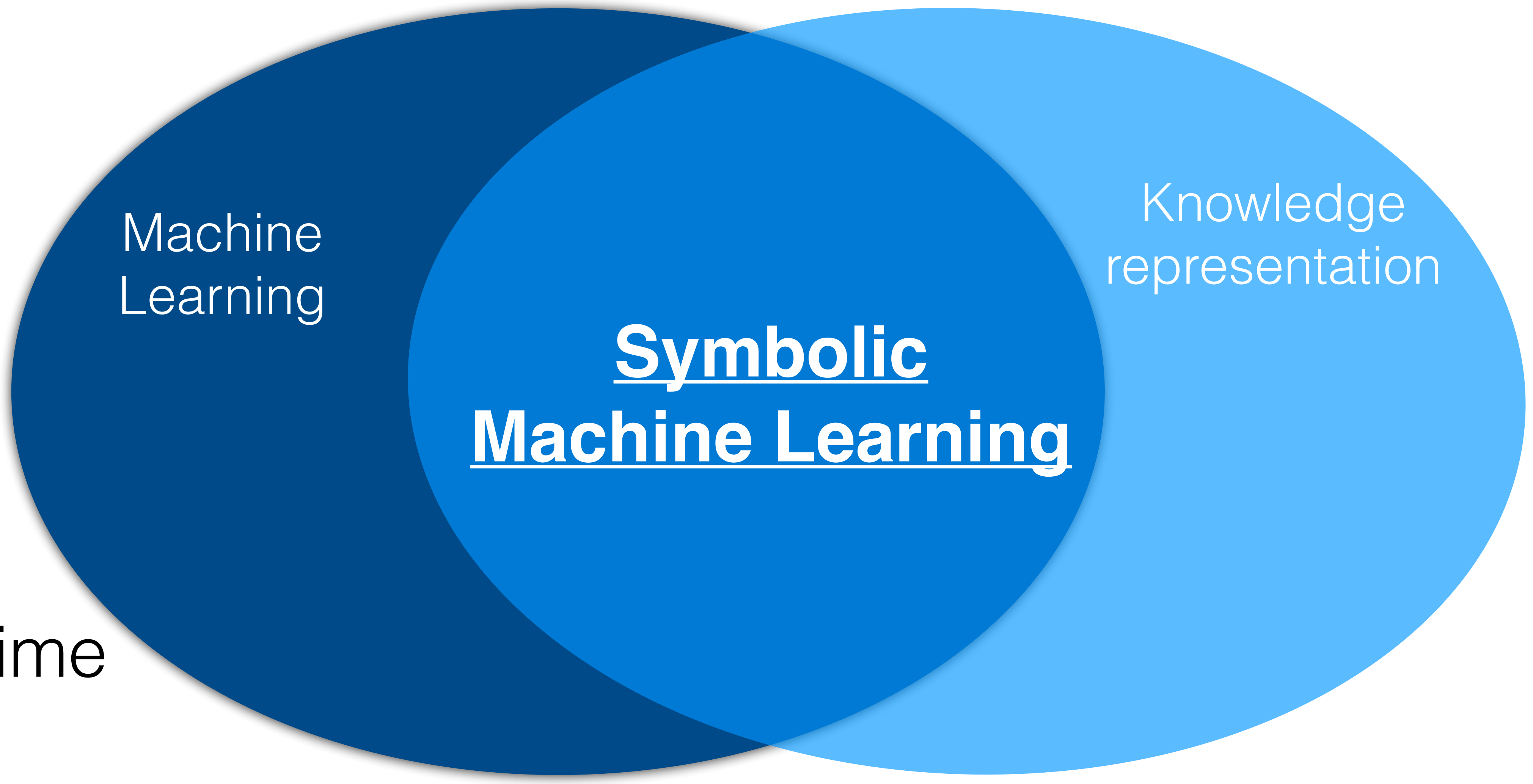
Ability to improve behaviour over time

Symbolic Machine Learning

Information extraction from data

Predictions about unseen data

Ability to improve behaviour over time



Human-readable representation of knowledge

Rigorous inference mechanisms for making predictions

Can be automatically verified

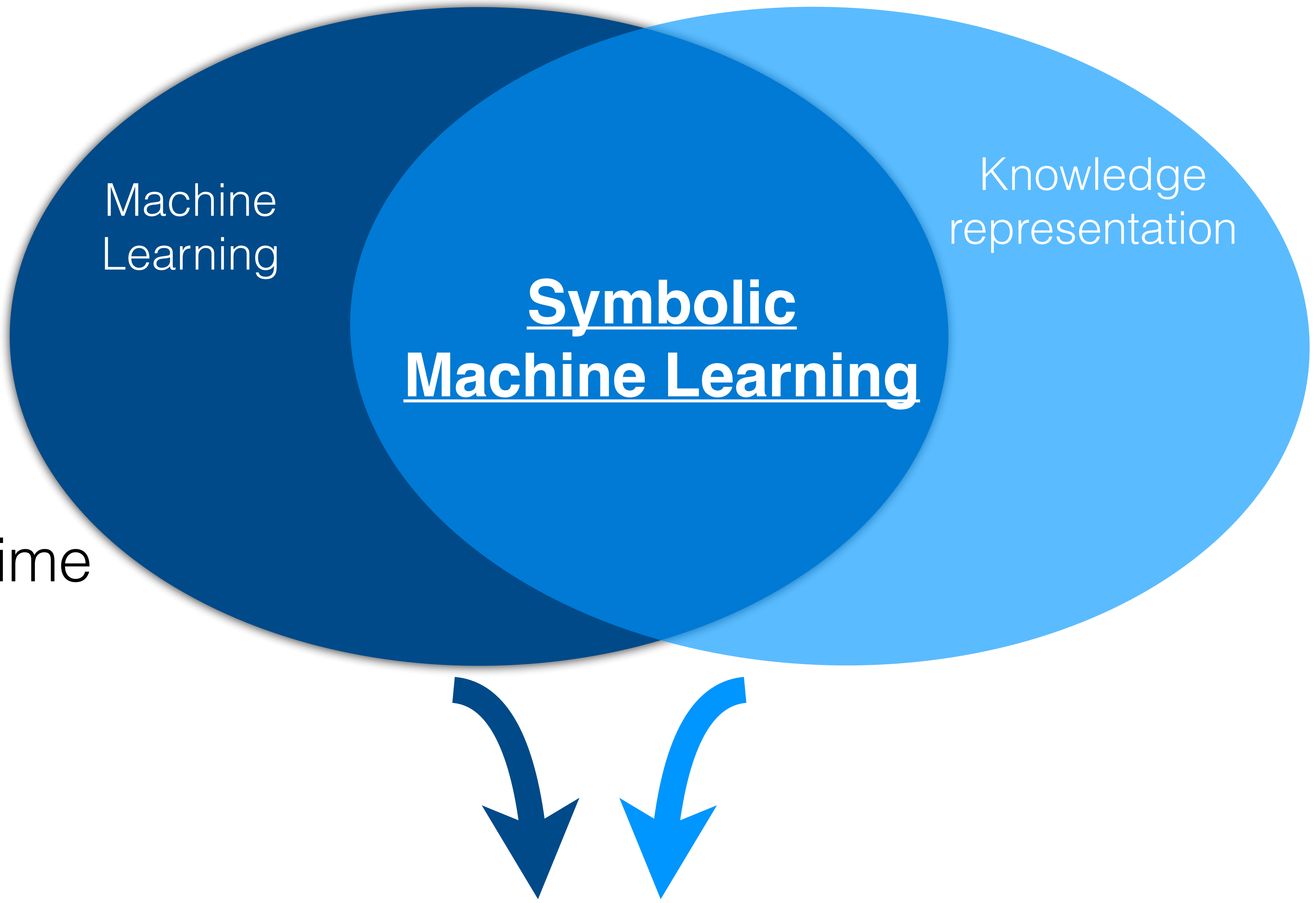
Imperial College London

Symbolic Machine Learning

Information extraction from data

Predictions about unseen data

Ability to improve behaviour over time



Human-readable representation of knowledge

Rigorous inference mechanisms for making predictions

Can be automatically verified

Symbolic Machine Learning can:

- learn from small amount of (noisy) dataset
- incorporate existing (partial) knowledge
- learn human-readable models
- support continuous and transfer learning

Imperial College London

Learning task: informal definition

A symbolic machine learning task is a T = <B, S_M, E^+, E^-> and a Covers relation over T

B	Background knowledge
S_M	Set of possible solutions
E^+	Set of positive examples
E^-	Set of negative examples

Learning task: informal definition

A symbolic machine learning task is a $T = <B, S_M, E^+, E^->$ and a Covers relation over T

B	Background knowledge
S_M	Set of possible solutions
E^+	Set of positive examples
E^-	Set of negative examples

The goal is to find a solution H in S_M that explains the given examples:

- Covers(B, H, e) for every $e \in E^+$
- ¬Covers(B, H, e) for every $e \in E^-$

Learning task: informal definition

A symbolic machine learning task is a $T = \langle B, S_M, E^+, E^- \rangle$ and a Covers relation over T

B	Background knowledge
S_M	Set of possible solutions
E^+	Set of positive examples
E^-	Set of negative examples

The goal is to find a solution H in S_M that explains the given examples:

- Covers(B, H, e) for every $e \in E^+$
- ¬Covers(B, H, e) for every $e \in E^-$

Different notions of Covers relation define different symbolic machine learning frameworks.

E.g.
- $B \cup H \models e$ for every $e \in E^+$
- $B \cup H \not\models e$ for every $e \in E^-$

Learning task: informal definition

A symbolic machine learning task is a $T = \langle B, S_M, E^+, E^- \rangle$ and a Covers relation over T

B	Background knowledge
S_M	Set of possible solutions
E^+	Set of positive examples
E^-	Set of negative examples

B	{ parent(ann, mary). parent(ann, tom). parent(tom, eve). parent(tom, ian). female(ann). female(mary). female(eve) }
S_M	{ daughter(X,Y) ← female(X). daughter(X,Y) ← parent(Y,X). daughter(X,Y) ← parent(Y,X), female(X) }
E^+	{ daughter(mary, ann). daughter(eve, tom) }
E^-	{ daughter(tom, ann). daughter(eve, ann) }

The goal is to find a solution H in S_M that explains the given examples:

- Covers(B, H, e) for every $e \in E^+$
- ¬Covers(B, H, e) for every $e \in E^-$

Different notions of Covers relation define different symbolic machine learning frameworks.

E.g.

- $B \cup H \models e$ for every $e \in E^+$
- $B \cup H \not\models e$ for every $e \in E^-$

Imperial College London

Learning task: informal definition

A symbolic machine learning task is a $T = \langle B, S_M, E^+, E^- \rangle$ and a Covers relation over T

B	Background knowledge
S_M	Set of possible solutions
E^+	Set of positive examples
E^-	Set of negative examples

B	{ parent(ann, mary). parent(ann, tom). parent(tom, eve). parent(tom, ian). female(ann). female(mary). female(eve) }
S_M	{ daughter(X,Y) ← female(X). daughter(X,Y) ← parent(Y,X). daughter(X,Y) ← parent(Y,X), female(X) }
E^+	{ daughter(mary, ann). daughter(eve, tom) }
E^-	{ daughter(tom, ann). daughter(eve, ann) }

The goal is to find a solution H in S_M that explains the given examples:

- Covers(B, H, e) for every $e \in E^+$
- ¬Covers(B, H, e) for every $e \in E^-$

Different notions of Covers relation define different symbolic machine learning frameworks.

E.g.
- $B \cup H \models e$ for every $e \in E^+$
- $B \cup H \not\models e$ for every $e \in E^-$

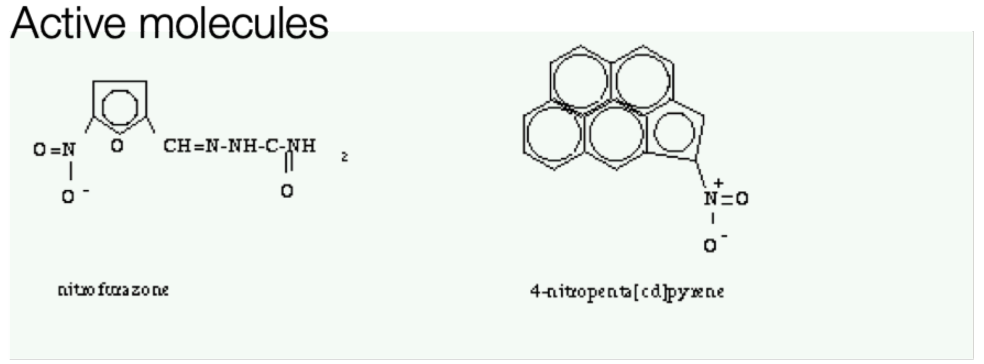
H = daughter(X,Y) ← parent(Y,X), female(X)

Early applications

- Predict mutagenicity of nitro compounds, relevant for prediction of carcinogenesis

Active molecules

E^+



Inactive molecules

E^-

B

active(f1).
atom(f1, $f1_1$, c, 21, 0.817).
atom(f1, $f1_2$, c, 21, -0.143).
atom(f1, $f1_3$, c, 21, -0.143).
……….
bond(f1, $f1_1$, $f1_2$, 7).
bond(f1, $f1_2$, $f1_3$, 7).
bond(f1, $f1_3$, $f1_4$, 7).
…………
logmutag(f1, 0.64).
lumo(f1, -1.785).
logp(f1, 1.01).

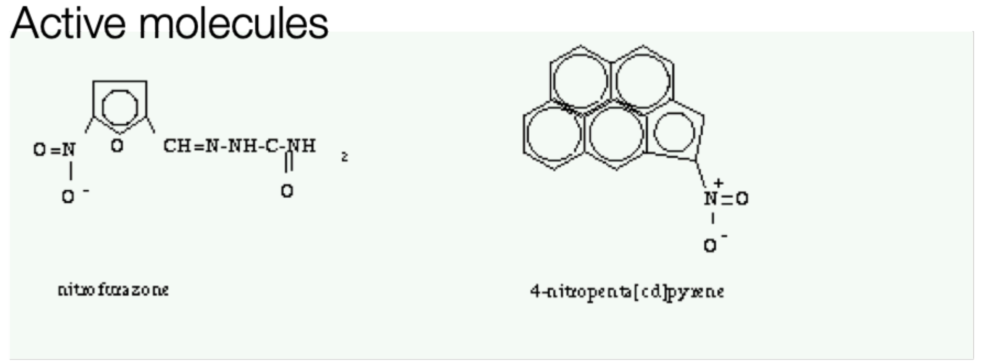
ring_size5(f1, [$f1_5$, $f1_1$, $f1_2$, $f1_3$, $f1_4$]).
……..

Early applications

- Predict mutagenicity of nitro compounds, relevant for prediction of carcinogenesis

E^+

Active molecules



E^-

Inactive molecules

B

active(f1).
atom(f1, $f1_1$, c, 21, 0.817).
atom(f1, $f1_2$, c, 21, -0.143).
atom(f1, $f1_3$, c, 21, -0.143).
………..
bond(f1, $f1_1$, $f1_2$, 7).
bond(f1, $f1_2$, $f1_3$, 7).
bond(f1, $f1_3$, $f1_4$, 7).
…………
logmutag(f1, 0.64).
lumo(f1, -1.785).
logp(f1, 1.01).

ring_size5(f1, [$f1_5$, $f1_1$, $f1_2$, $f1_3$, $f1_4$]).
……..

H

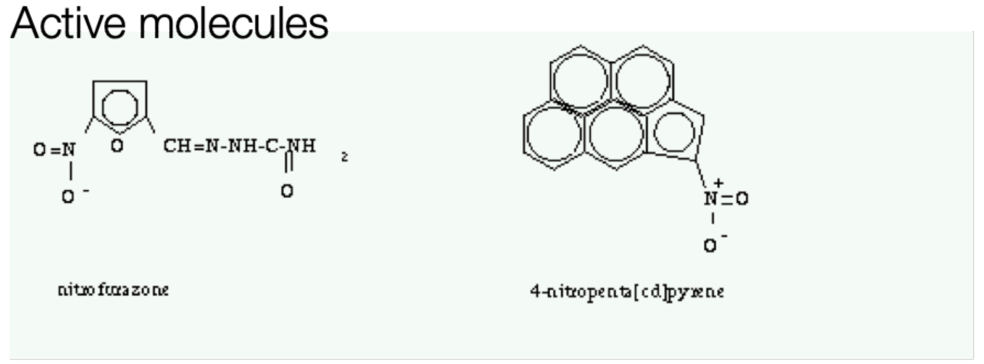
mutagenic(M) ← ring_size5(M, L),
atom(M,A1, _,_,_),
atom (M,A2,_,_,_),
member(A1,L),
bond(M,A1,A2, 2)

Early applications

- Predict mutagenicity of nitro compounds, relevant for prediction of carcinogenesis

Active molecules

E^+



Inactive molecules

E^-

B

active(f1).
atom(f1, $f1_1$, c, 21, 0.817).
atom(f1, $f1_2$, c, 21, -0.143).
atom(f1, $f1_3$, c, 21, -0.143).
……….
bond(f1, $f1_1$, $f1_2$, 7).
bond(f1, $f1_2$, $f1_3$, 7).
bond(f1, $f1_3$, $f1_4$, 7).
…………
logmutag(f1, 0.64).
lumo(f1, -1.785).
logp(f1, 1.01).

ring_size5(f1, [$f1_5$, $f1_1$, $f1_2$, $f1_3$, $f1_4$]).
……..

H

mutagenic(M) ← ring_size5(M, L),
atom(M,A1, _,_,_),
atom (M,A2,_,_,_),
member(A1,L),
bond(M,A1,A2, 2)

- Learn regular grammars, from observations of positive and negative example strings

$_0$ She $_1$ ran $_2$ quickly $_3$

E^+ = s(0,3)

B

np(X,Y)←word("She",X,Y).
mod(X,Y)←word("quickly",X,Y).
s(X,Y)←np(X,Z), vp(Z,Y).
vp(X,Y)←v(X,Y).

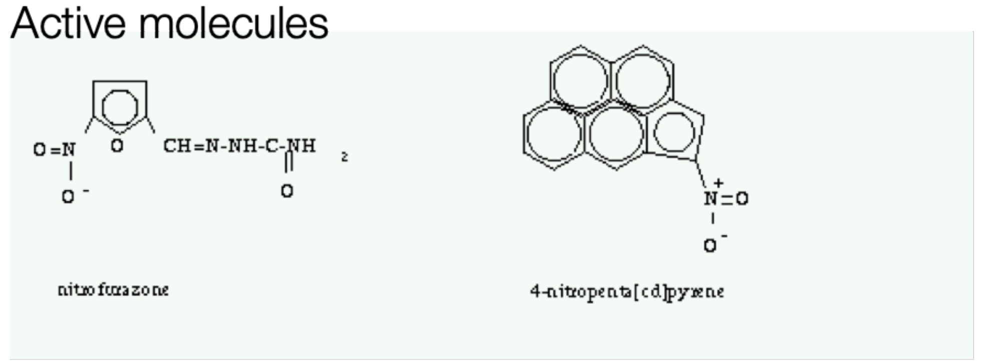
word("She",0,1)
word("quickly", 2,3)
word(ran,1,2)
←v(1,3)

Early applications

- Predict mutagenicity of nitro compounds, relevant for prediction of carcinogenesis

E^+

Active molecules



E^-

Inactive molecules

B

active(f1).
atom(f1, $f1_1$, c, 21, 0.817).
atom(f1, $f1_2$, c, 21, -0.143).
atom(f1, $f1_3$, c, 21, -0.143).
………..
bond(f1, $f1_1$, $f1_2$, 7).
bond(f1, $f1_2$, $f1_3$, 7).
bond(f1, $f1_3$, $f1_4$, 7).
…………
logmutag(f1, 0.64).
lumo(f1, -1.785).
logp(f1, 1.01).

ring_size5(f1, [$f1_5$, $f1_1$, $f1_2$, $f1_3$, $f1_4$]).
……..

H

mutagenic(M) ← ring_size5(M, L),
atom(M,A1, _,_,_),
atom (M,A2,_,_,_),
member(A1,L),
bond(M,A1,A2, 2)

- Learn regular grammars, from observations of positive and negative example strings

$_0$ She $_1$ ran $_2$ quickly $_3$

E^+ = s(0,3)

B

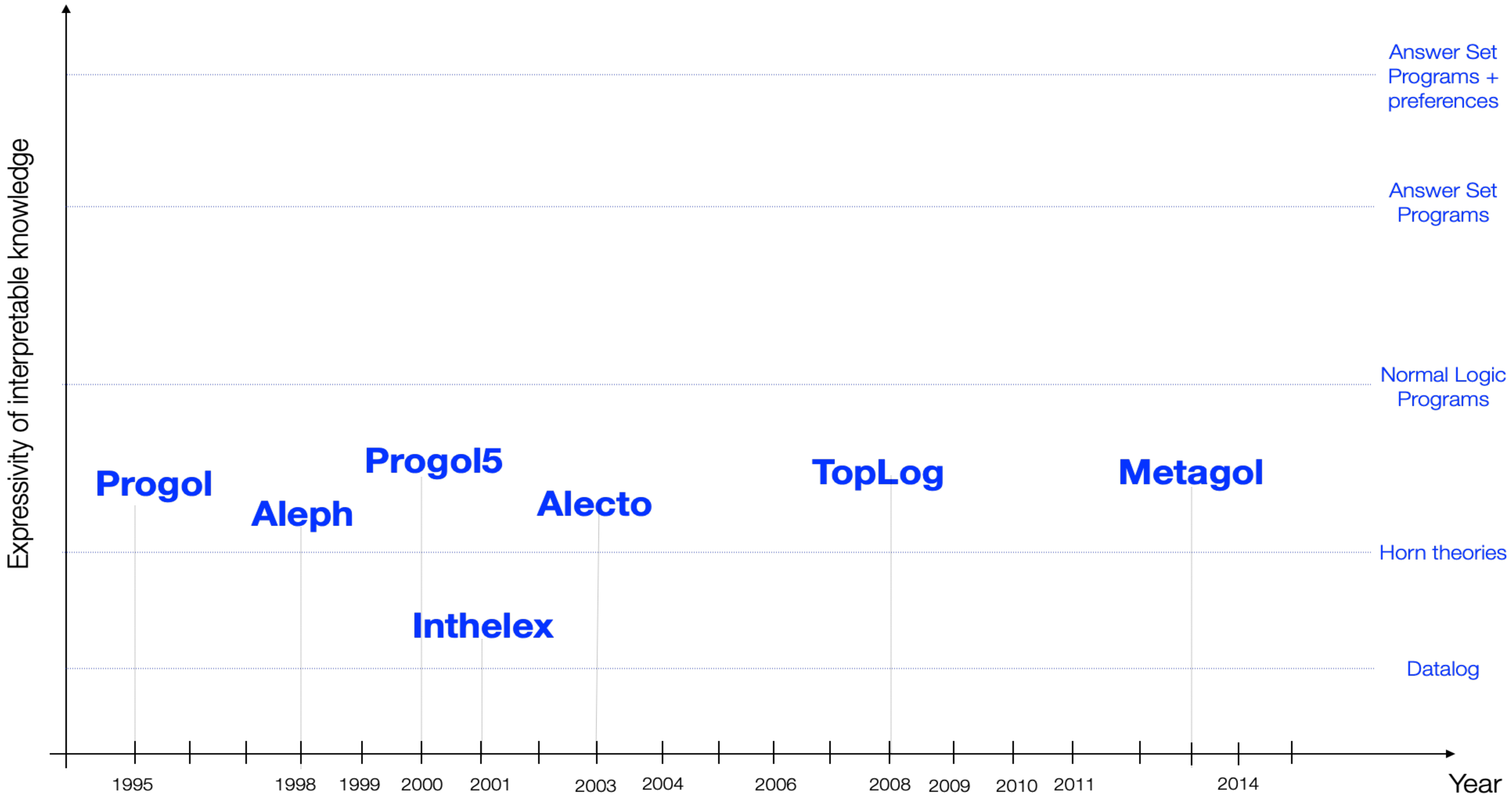
np(X,Y)←word("She",X,Y).
mod(X,Y)←word("quickly",X,Y).
s(X,Y)←np(X,Z), vp(Z,Y).
vp(X,Y)←v(X,Y).

word("She",0,1)
word("quickly", 2,3)
word(ran,1,2)
←v(1,3)

H

v(1, 2)
vp(Start, End) ← vp(Start, Middle),
mod(Middle, End)

Early algorithms and systems

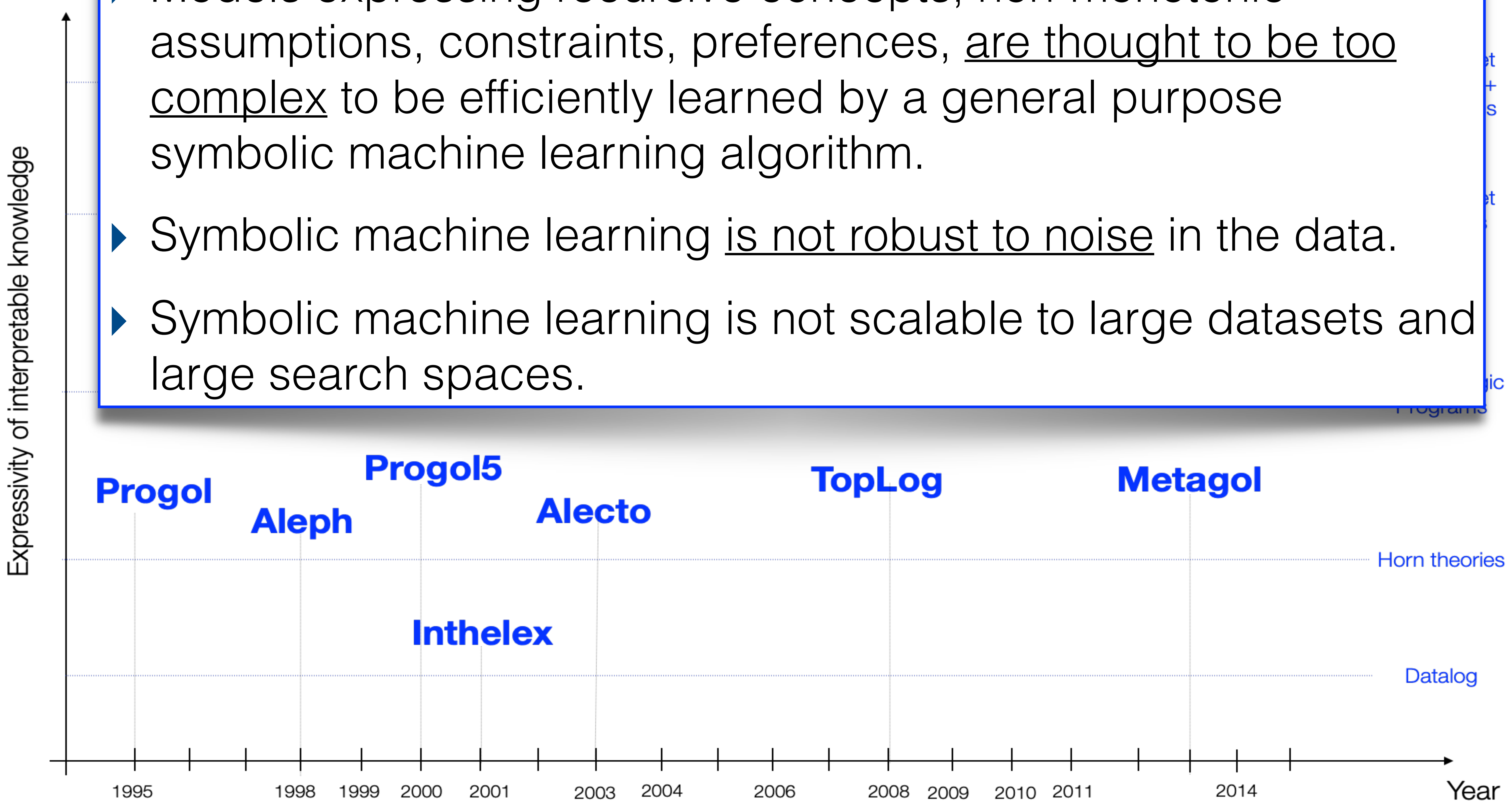


Imperial College
London

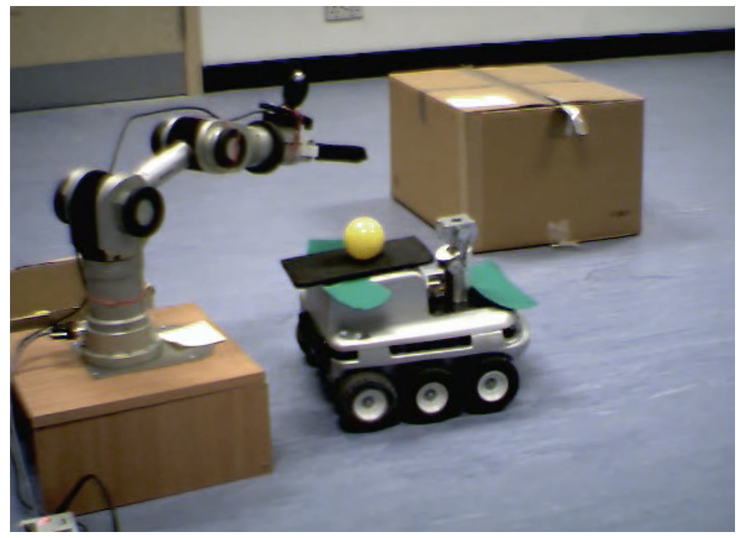
Early algorithms and systems

Three Main Misconceptions:

- Models expressing recursive concepts, non-monotonic assumptions, constraints, preferences, are thought to be too complex to be efficiently learned by a general purpose symbolic machine learning algorithm.
- Symbolic machine learning is not robust to noise in the data.
- Symbolic machine learning is not scalable to large datasets and large search spaces.



Learning complex but interpretable models



- Behaving autonomously in the real-world requires learning default assumptions

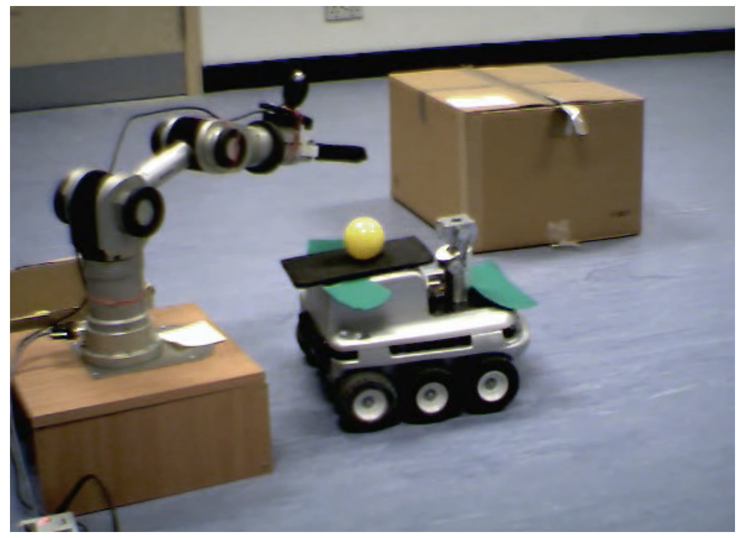
  succeeds(putdown, T) ← not happened(move(loc1,loc2),T-2)

Imperial College London

Learning complex but interpretable models

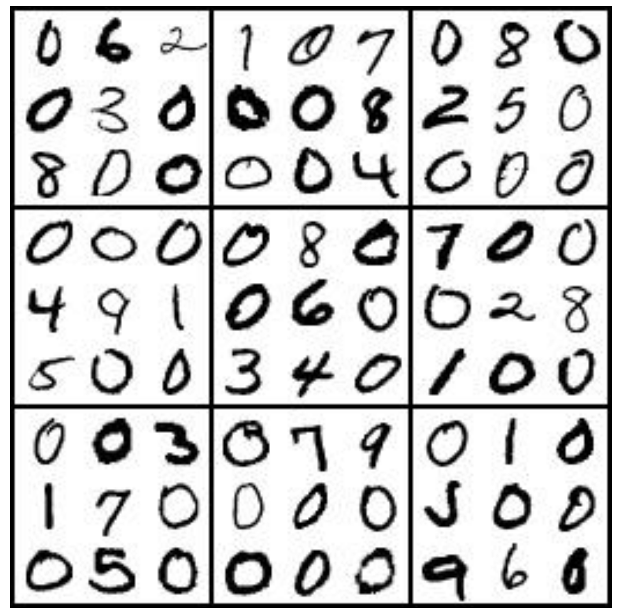
- Behaving autonomously in the real-world requires learning default assumptions

  succeeds(putdown, T) ← not happened(move(loc1,loc2),T-2)



- Guaranteeing correct and safe decisions require learning constraints

  falsity ← value(V, C1), value(V, C2), same_col(C1, C2).
  falsity ← value(V, C1), value(V, C2), same_row(C1, C2).
  falsity ← value(V, C1), value(V, C2), same_block(C1, C2).

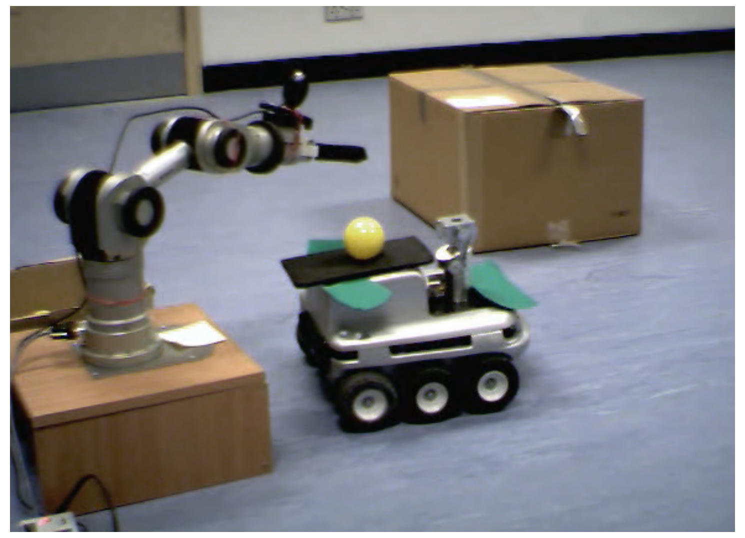


Imperial College London

Learning complex but interpretable models

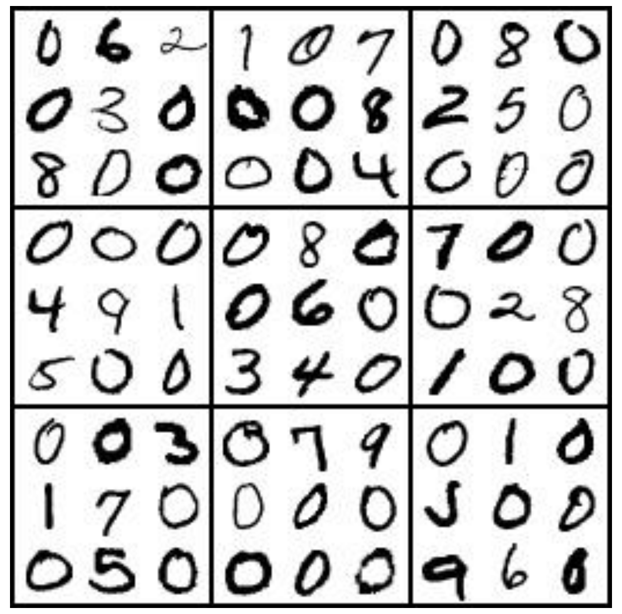
- Behaving autonomously in the real-world requires learning default assumptions

  succeeds(putdown, T) ← not happened(move(loc1,loc2),T-2)



- Guaranteeing correct and safe decisions require learning constraints

  falsity ← value(V, C1), value(V, C2), same_col(C1, C2).
  falsity ← value(V, C1), value(V, C2), same_row(C1, C2).
  falsity ← value(V, C1), value(V, C2), same_block(C1, C2).

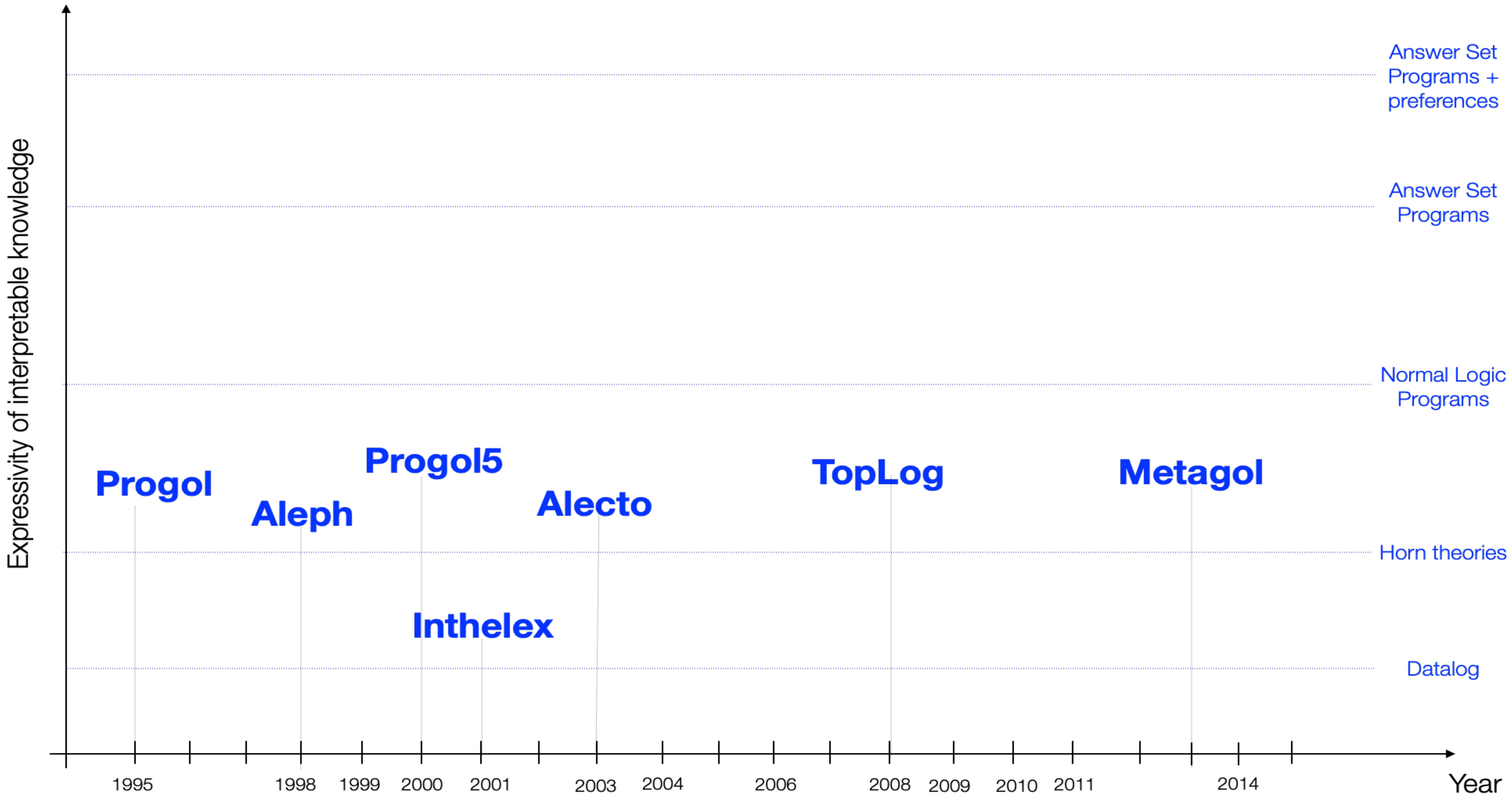


- Assisting humans in their decision making require learning their preferences

  :~ mode(Zone, walk), crime_rating(Zone, R), R > 4.[1@3]
  :~ mode(Zone, bus).[1@2]
  :~ mode(Zone, walk), distance(Zone, D).[D@1]

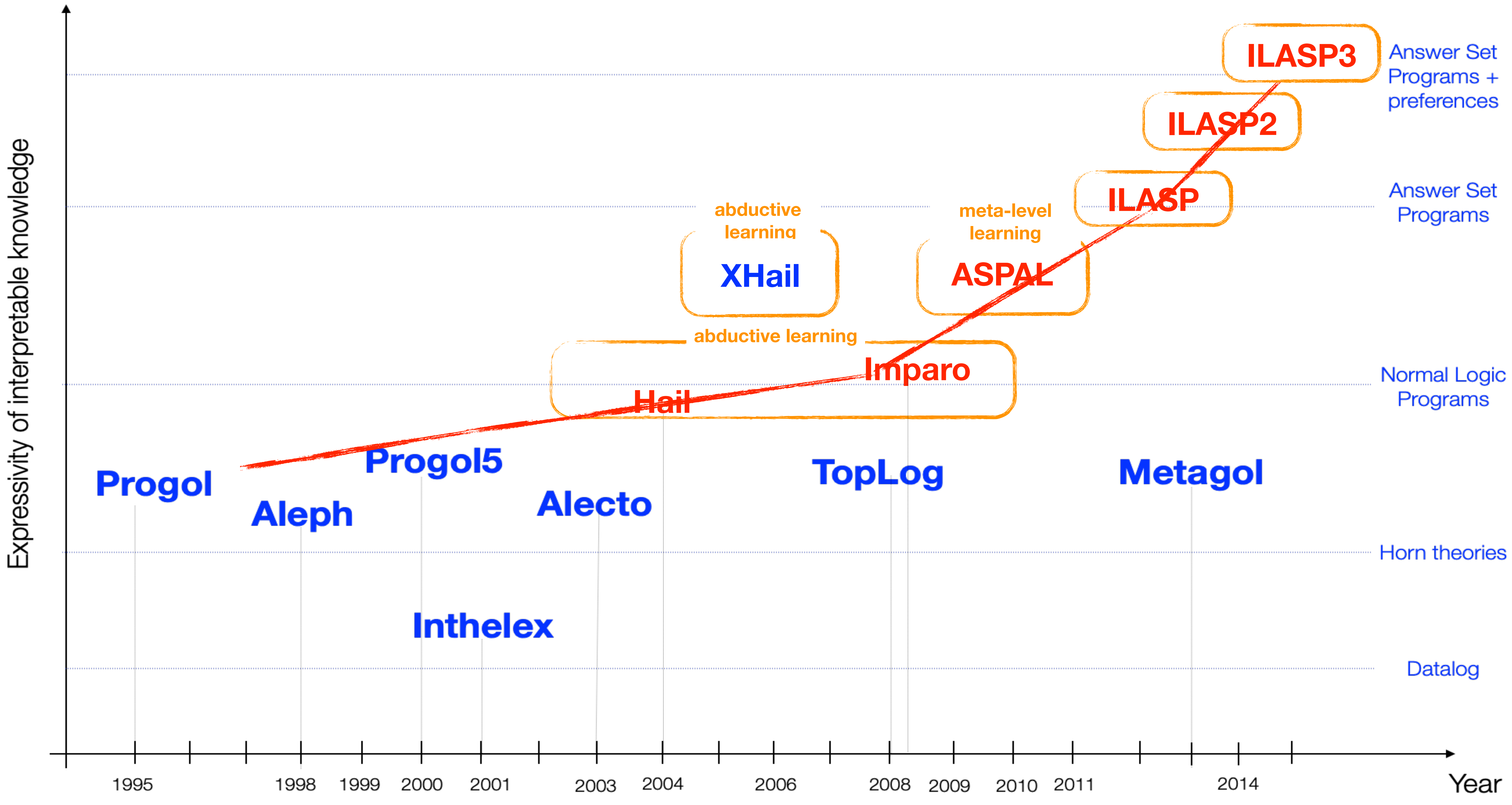
Imperial College London

…Our recent advancements



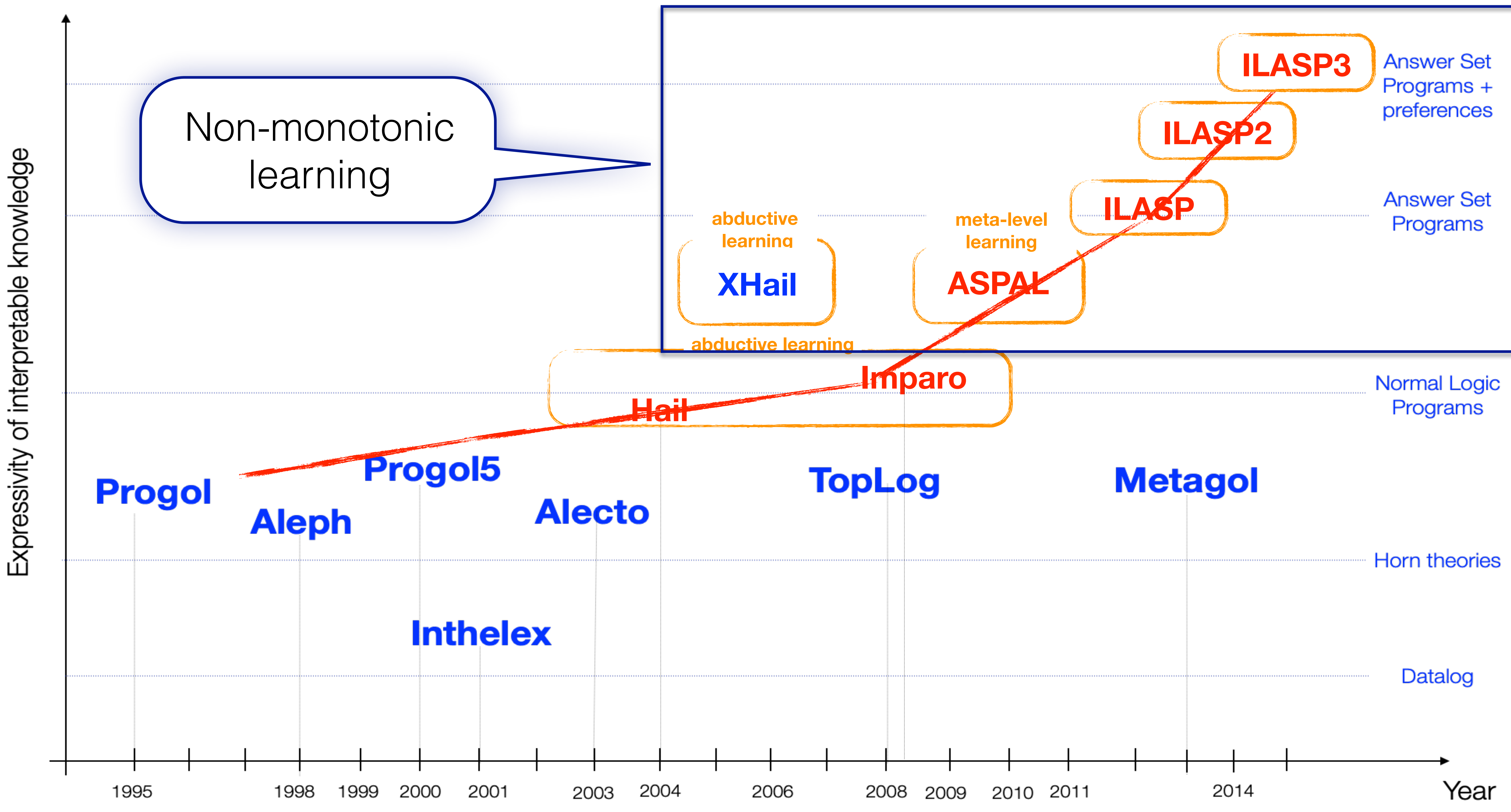
Imperial College London

…Our recent advancements

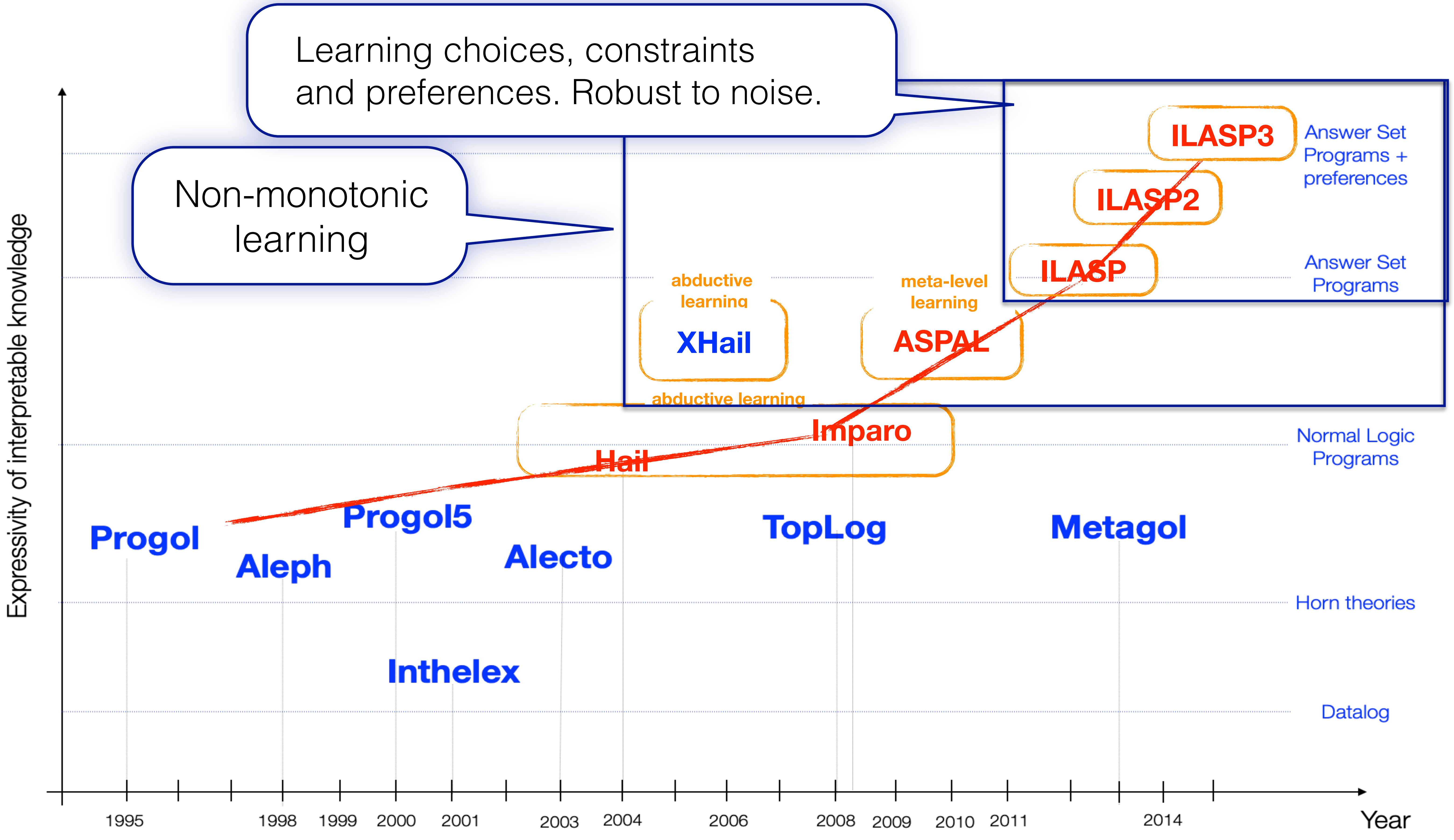


Imperial College London

…Our recent advancements



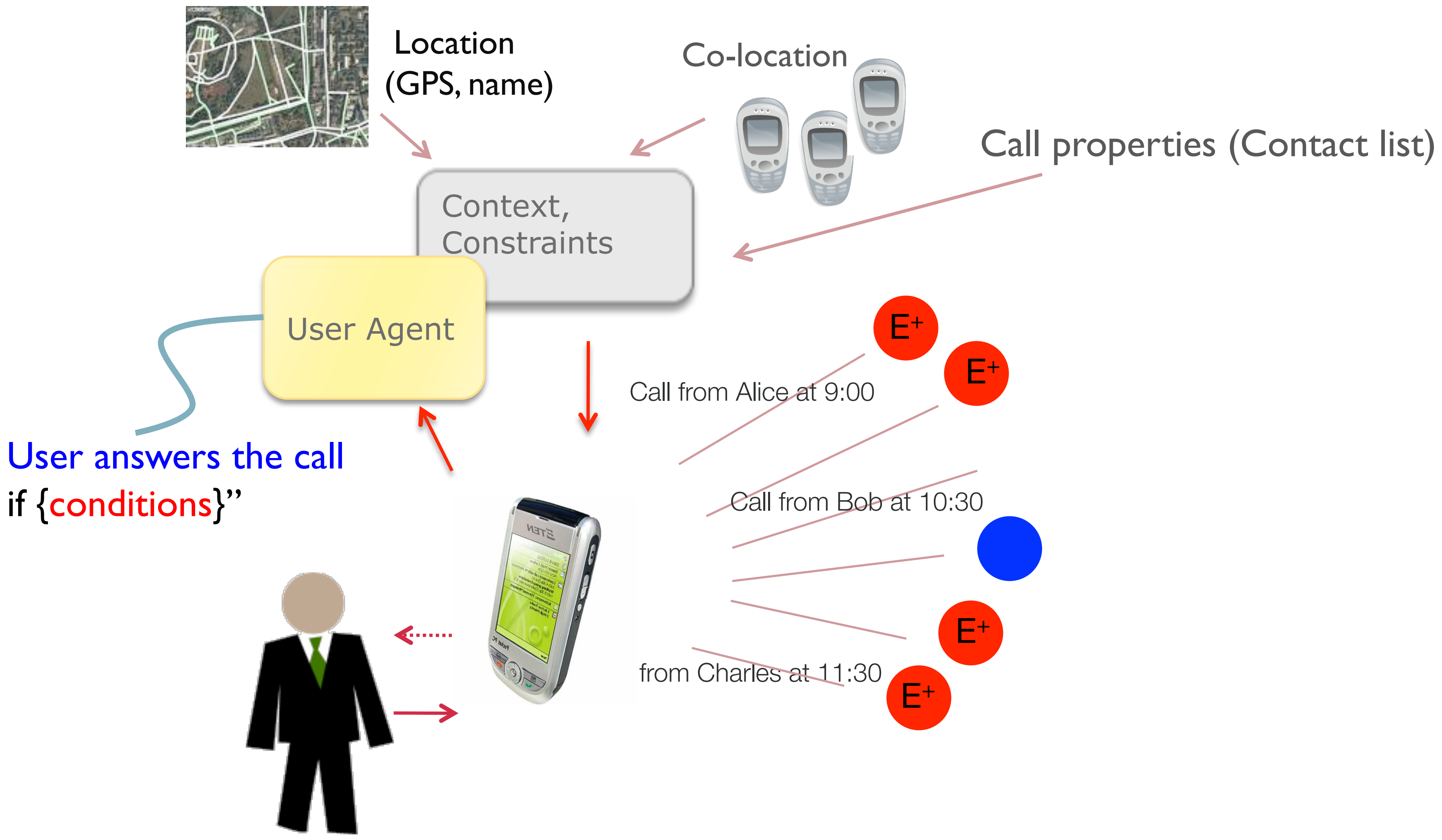
…Our recent advancements



Learning Rule-based Policies

Learning user behaviour models in pervasive systems

- Devices are able to continuously learn policies from (user) past actions
- Learned policies are used to automatically adapt their behaviours and reduce human intervention



D. Corapi, A. Russo, E. Lupu, Inductive Logic Programming in Answer Set Programming. ILP 2011

Imperial College London

Learning Rule-based Policies

Learning user behaviour models in pervasive systems

- Devices are able to continuously learn policies from (user) past actions
- Learned policies are used to automatically adapt their behaviours and reduce human intervention

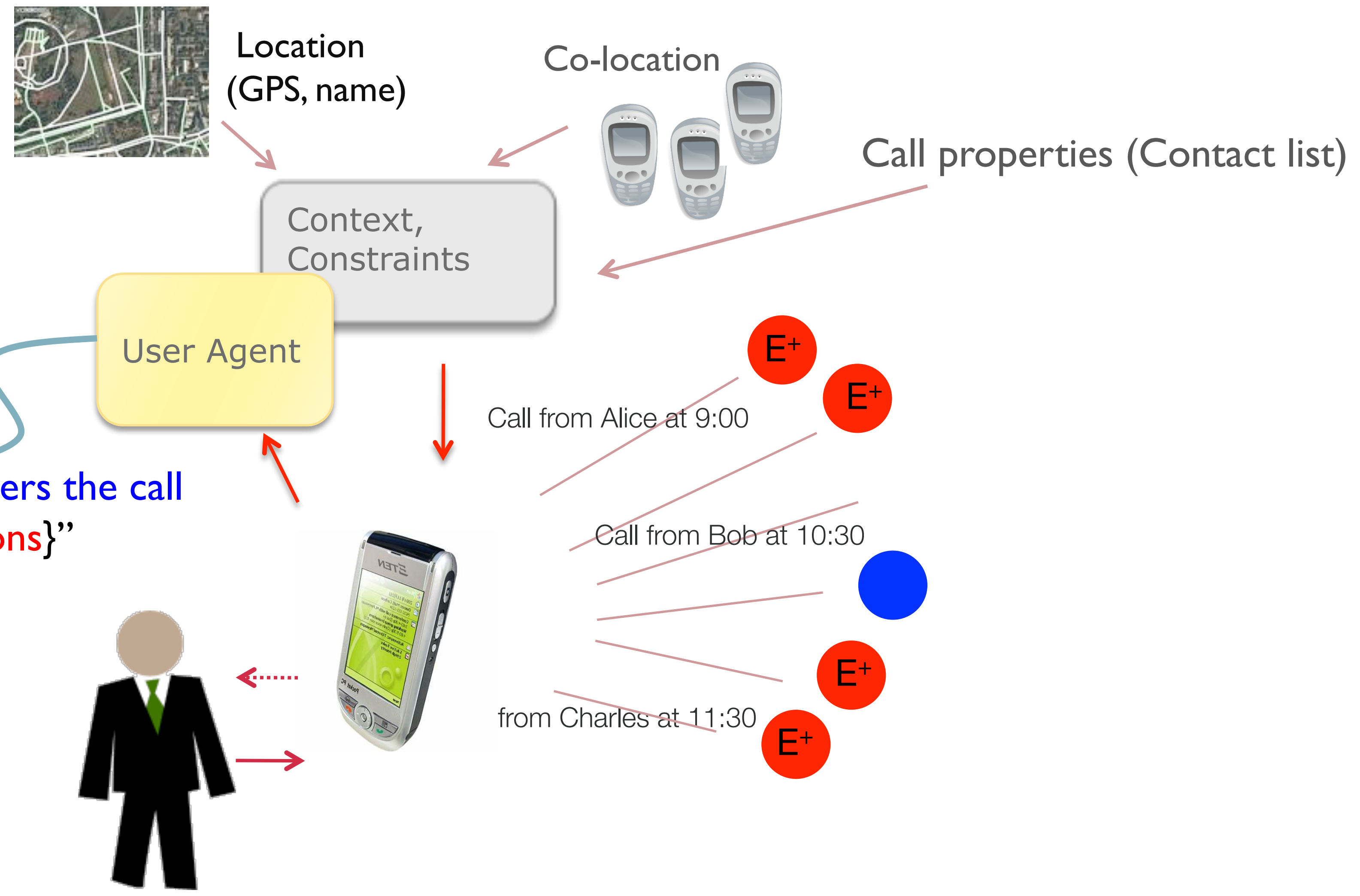
answerCall(…) $\leftarrow$ $condition_1$, …, $condition_{maxC}$

……

answerCall(…) $\leftarrow$ $condition_{1,maxR}$, …, $condition_{maxC,maxR}$

=(+volume, #volume).
=(+vibrator, #vibrator).
=(+battery_level, #battery_level).
=(+screen_brightness, #screen_brightness).
=(+headset, #headset).
=(+screen_status, #screen_status).
=(+light_level, #light_level).
=(+battery_charging, #battery_charging).
weekday(+date)).
weekend(+date)).
evening(+time)).
morning(+time)).
afternoon(+time)).
in_call(+date, +time)).
at(+date, +time, #cell)).
nearDevice(+date, +time, #device)).
neighbourhood(+cell, #cell)).
user_been_in(+date, +t
(user_is_active(+date, +time)).
phone_charging(+date, +time)).
phone_on(+date, +time)).
user_is_using_app(+date, +time, #app)).
time_before_h(+time, #hour).
(time_after_h(+time, #hour).

Around 5000 choices for conditions



D. Corapi, A. Russo, E. Lupu, Inductive Logic Programming in Answer Set Programming. ILP 2011

Learning Rule-based Policies

Learning user behaviour models in pervasive systems

- Devices are able to continuously learn policies from (user) past actions
- Learned policies are used to automatically adapt their behaviours and reduce human intervention

$answerCall(\ldots) \leftarrow condition_{1}, \ldots, condition_{maxC}$

......

$answerCall(\ldots) \leftarrow condition_{1,maxR}, \ldots, condition_{maxC,maxR}$

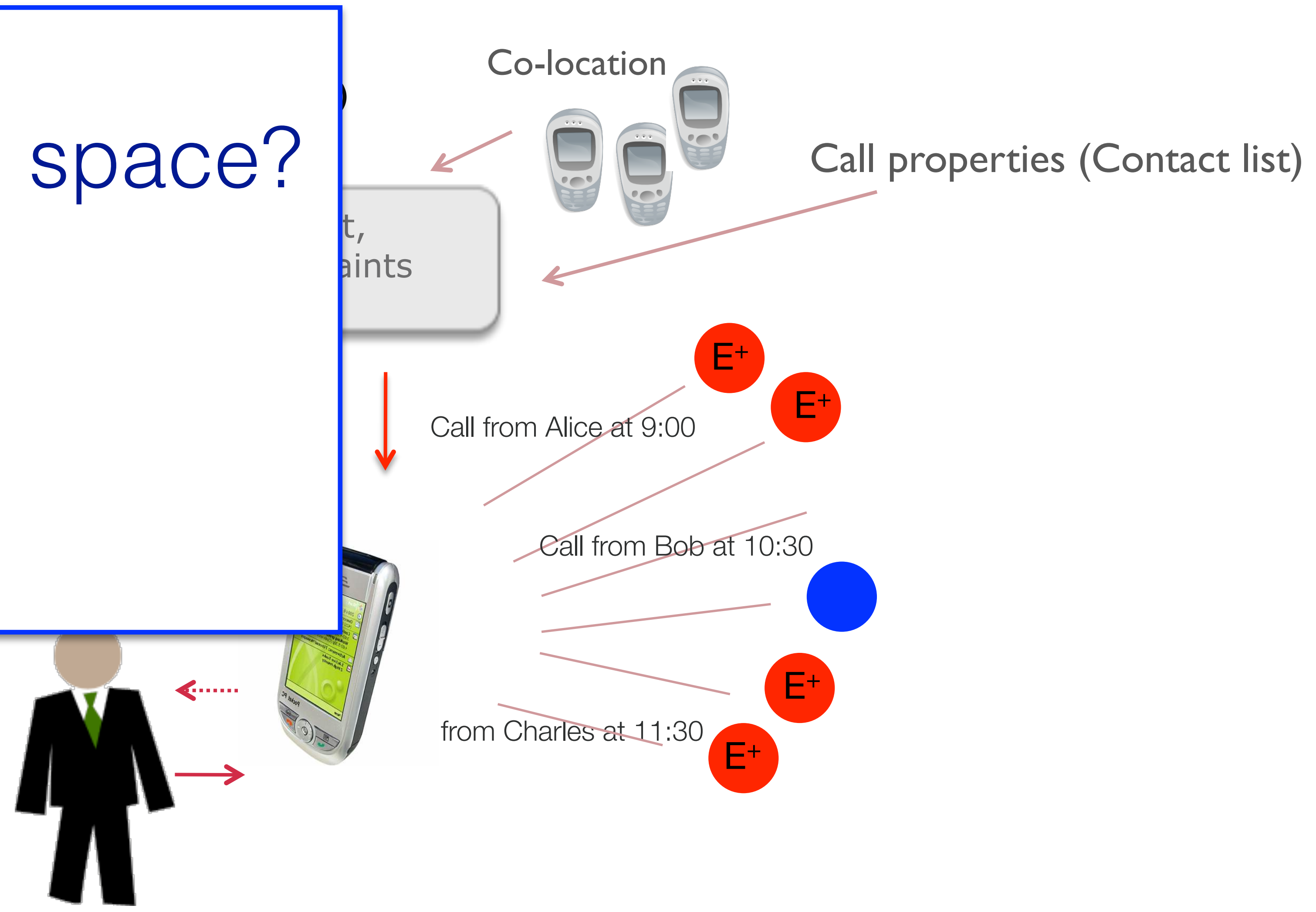
=(+volume, #volu
=(+vibrator, #vibr
=(+battery_level,
=(+screen_bright
=(+headset, #hea
=(+screen_status
=(+light_level, #li
=(+battery_charg
weekday(+date))
weekend(+date))
evening(+time)).
morning(+time)).
afternoon(+time)
in_call(+date, +ti
at(+date, +time, #cell).
nearDevice(+date, +time, #device)).
neighbourhood(+cell, #cell)).
user_been_in(+date, +t
(user_is_active(+date, +time)).
phone_charging(+date, +time)).
phone_on(+date, +time)).
user_is_using_app(+date, +time, #app)).
time_before_h(+time, #hour).
(time_after_h(+time, #hour).

choices for conditions

How many policies in the search space?

$((5000)^{maxC})^{maxR}$

$5000^{4*10} \cong 10^{147}$ possible



D. Corapi, A. Russo, E. Lupu, Inductive Logic Programming in Answer Set Programming. ILP 2011

Learning Rule-based Policies

Learning user behaviour models in pervasive systems

- Devices are able to continuously learn policies from (user) past actions
- Learned policies are used to automatically adapt their behaviours and reduce human intervention

ASPAL

answerCall(…) ← $condition_1$, …, $condition_{maxC}$

……

answerCall(…) ← $condition_{1,maxR}$, …, $condition_{maxC,maxR}$

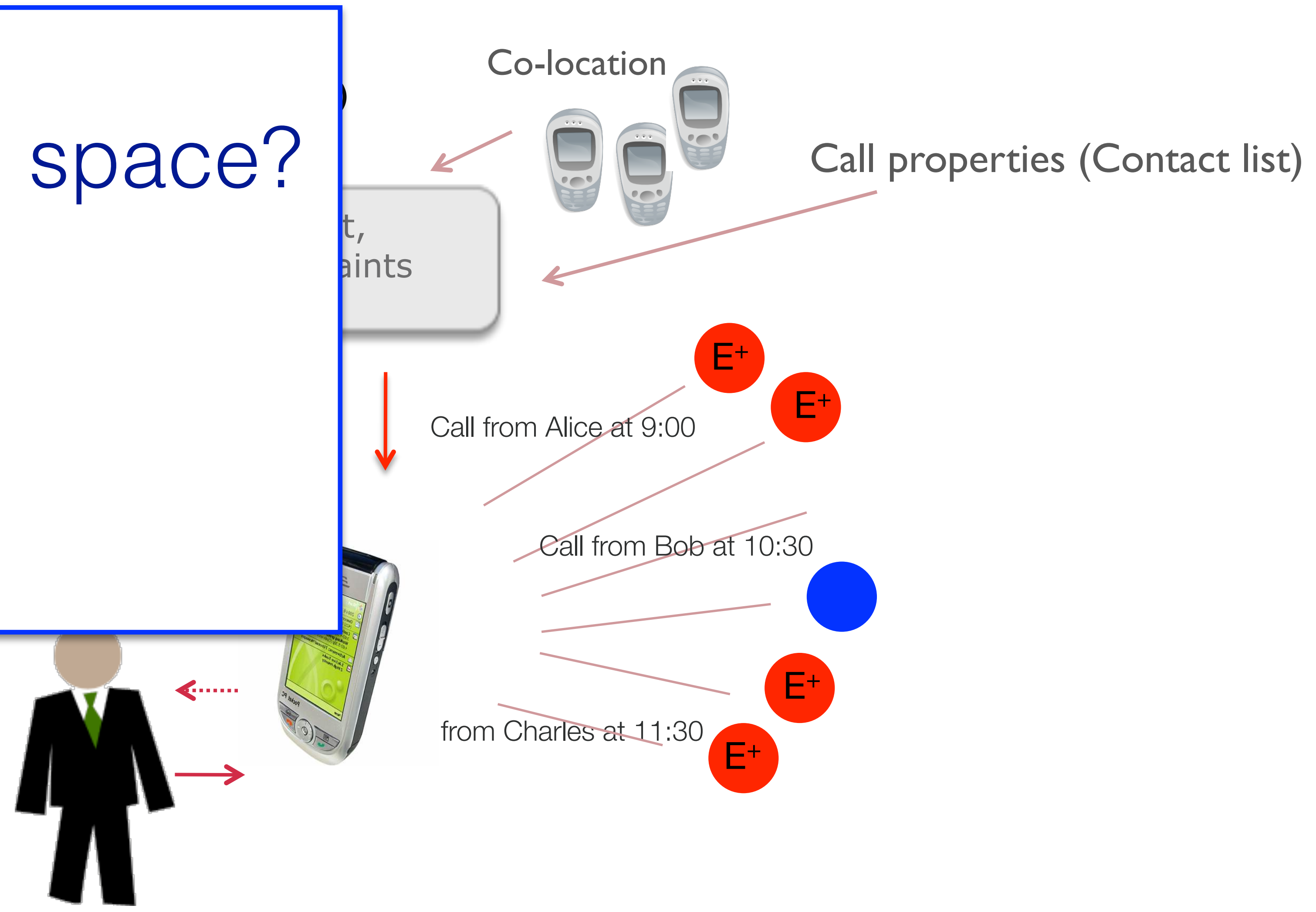
=(+volume, #volu
=(+vibrator, #vibr
=(+battery_level,
=(+screen_bright
=(+headset, #hea
=(+screen_status
=(+light_level, #li
=(+battery_charg
weekday(+date)).
weekend(+date)).
evening(+time)).
morning(+time)).
afternoon(+time)
in_call(+date, +ti
at(+date, +time, #cell).
nearDevice(+date, +time, #device)).
neighbourhood(+cell, #cell)).
user_been_in(+date, +t
(user_is_active(+date, +time)).
phone_charging(+date, +time)).
phone_on(+date, +time)).
user_is_using_app(+date, +time, #app)).
time_before_h(+time, #hour).
(time_after_h(+time, #hour).

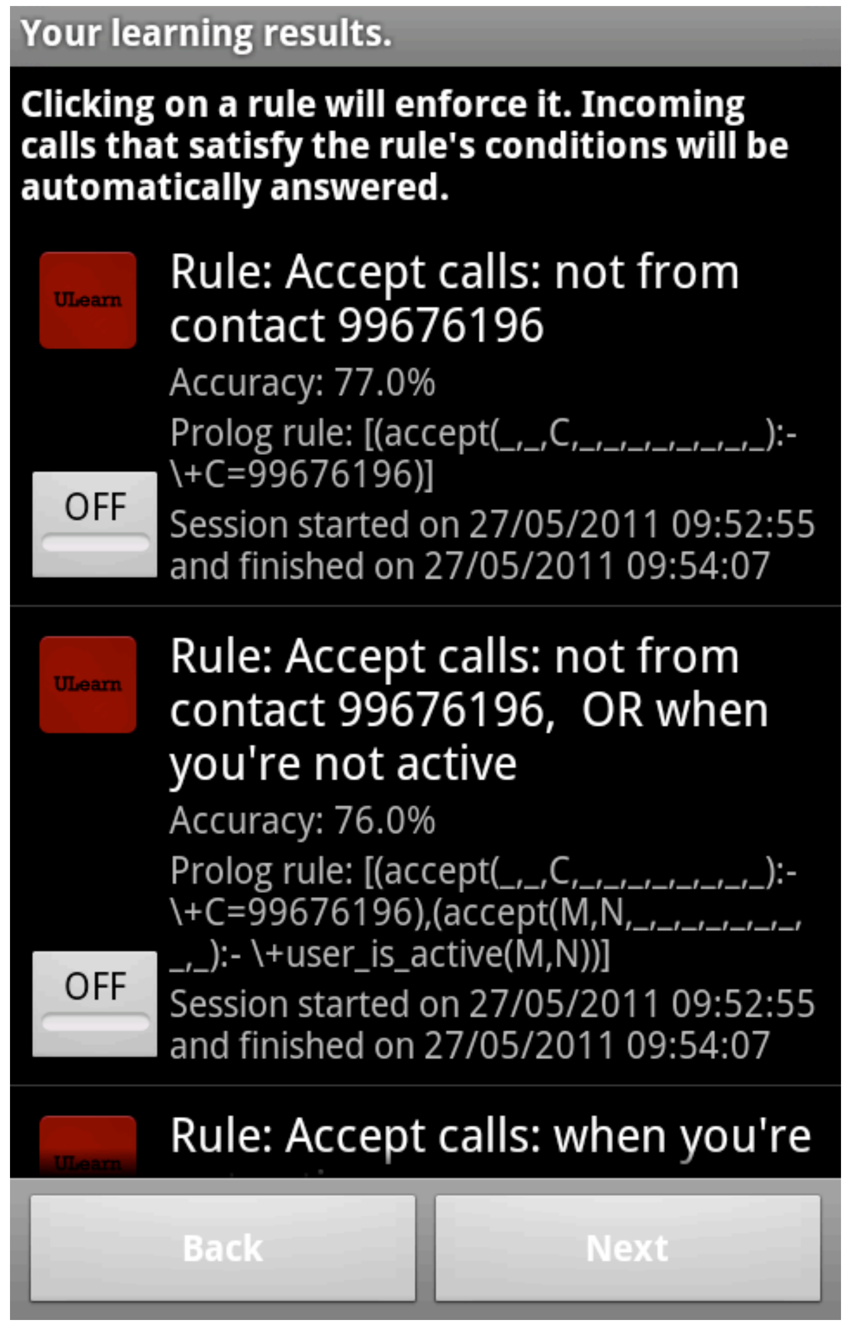
choices for conditions

How many policies in the search space?

$((5000)^{maxC})^{maxR}$

$5000^{4*10} \cong 10^{147}$ possible





D. Corapi, A. Russo, E. Lupu, Inductive Logic Programming in Answer Set Programming. ILP 2011

Learning Rule-based Policies

Learning user behaviour models in pervasive systems

- Devices are able to continuously learn policies from (user) past actions
- Learned policies are used to automatically adapt their behaviours and reduce human intervention

ASPAL

answerCall(…) ← $condition_1, \ldots, condition_{maxC}$

……

answerCall(…) ← $condition_{1,maxR}, \ldots, condition_{maxC,maxR}$

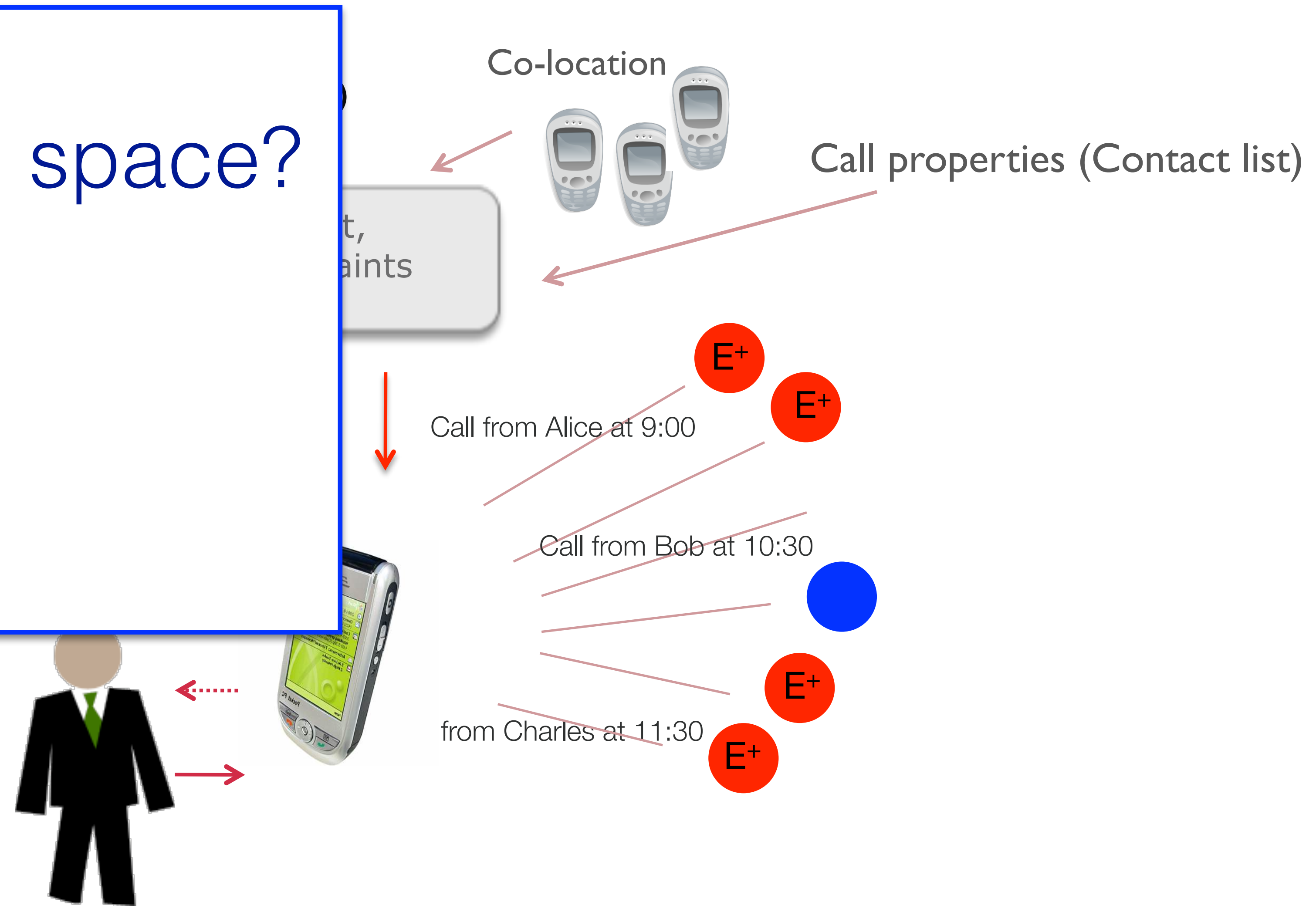
=(+volume, #volu
=(+vibrator, #vibr
=(+battery_level,
=(+screen_bright
=(+headset, #hea
=(+screen_status
=(+light_level, #li
=(+battery_charg
weekday(+date)).
weekend(+date)).
evening(+time)).
morning(+time)).
afternoon(+time)
in_call(+date, +ti
at(+date, +time, #cell).
nearDevice(+date, +time, #device)).
neighbourhood(+cell, #cell)).
user_been_in(+date, +t
(user_is_active(+date, +time)).
phone_charging(+date, +time)).
phone_on(+date, +time)).
user_is_using_app(+date, +time, #app)).
time_before_h(+time, #hour).
(time_after_h(+time, #hour).

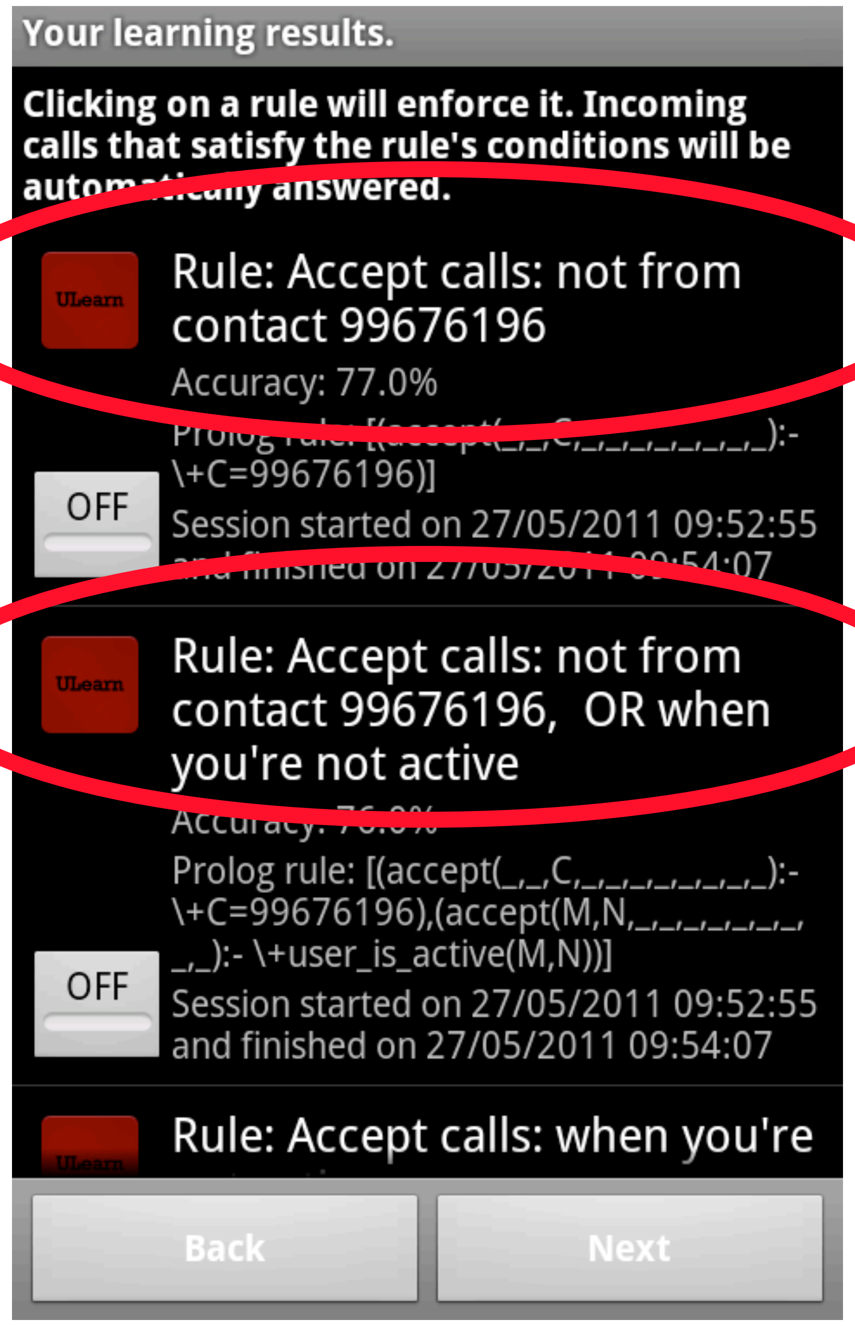
choices for conditions

How many policies in the search space?

$((5000)^{maxC})^{maxR}$

$5000^{4*10} \cong 10^{147}$ possible





D. Corapi, A. Russo, E. Lupu, Inductive Logic Programming in Answer Set Programming. ILP 2011

Inductive Learning of Answer Set Programs

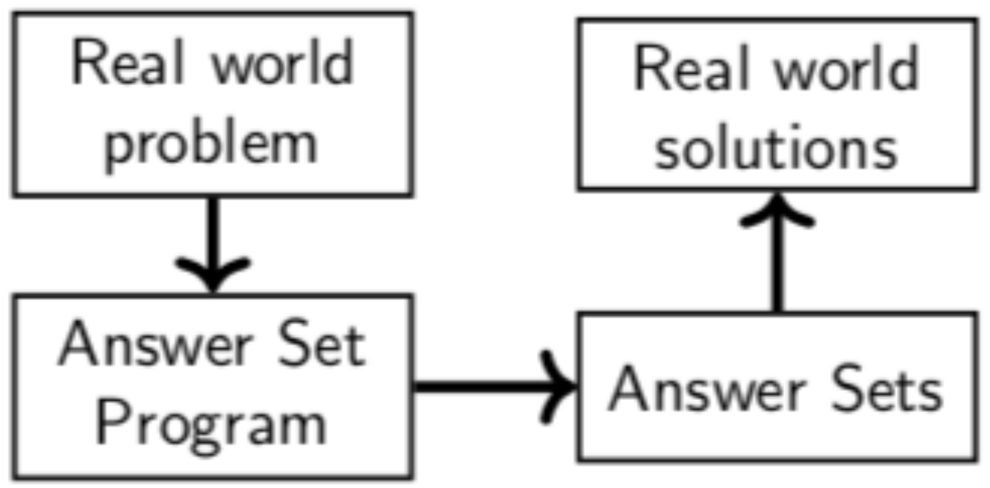
The ILASP Systems

Inductive Learning of Answer Set Programs

The ILASP Systems

Answer Set Programming

Expressive Declarative Environment
for Reasoning Logically

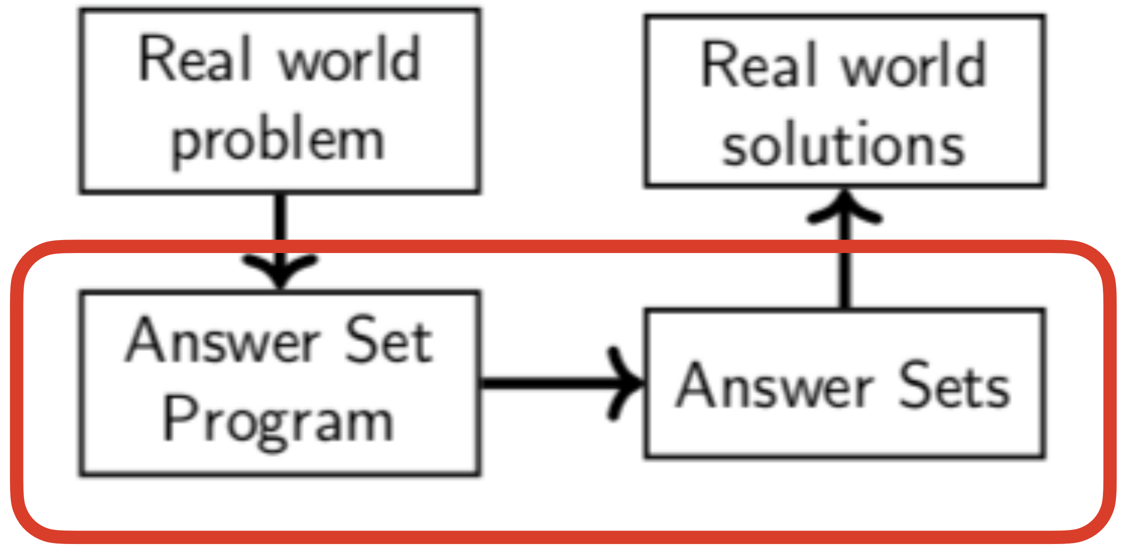


Inductive Learning of Answer Set Programs

The ILASP Systems

Answer Set Programming

Expressive Declarative Environment
for Reasoning Logically



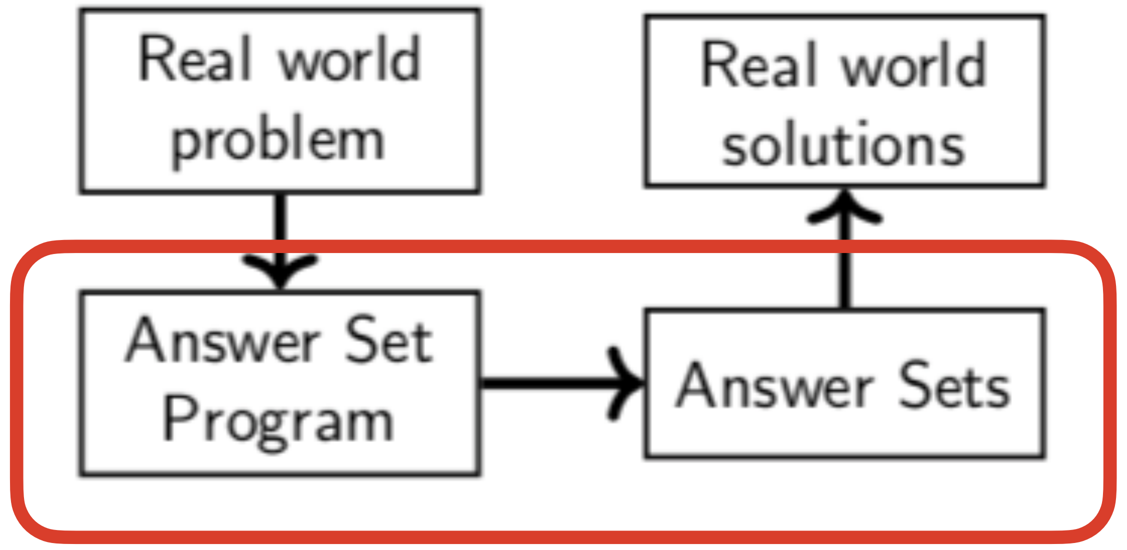
Imperial College London

Inductive Learning of Answer Set Programs

The ILASP Systems

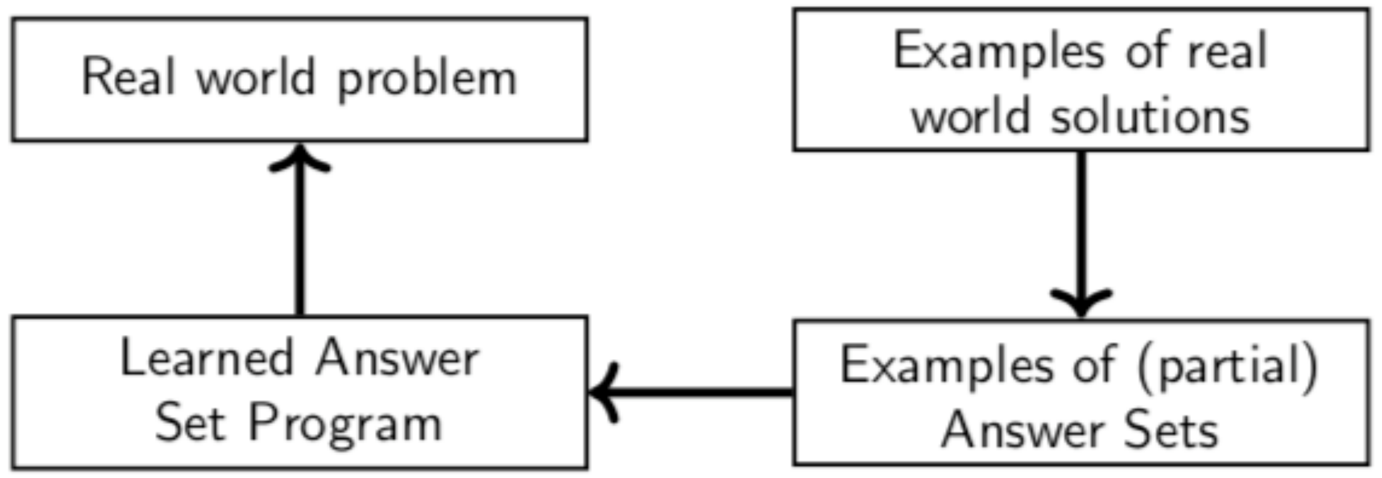
Answer Set Programming

Expressive Declarative Environment
for Reasoning Logically



ILASP

Expressive Declarative Environment
for Learning Logically



ILASP
Learning Logically

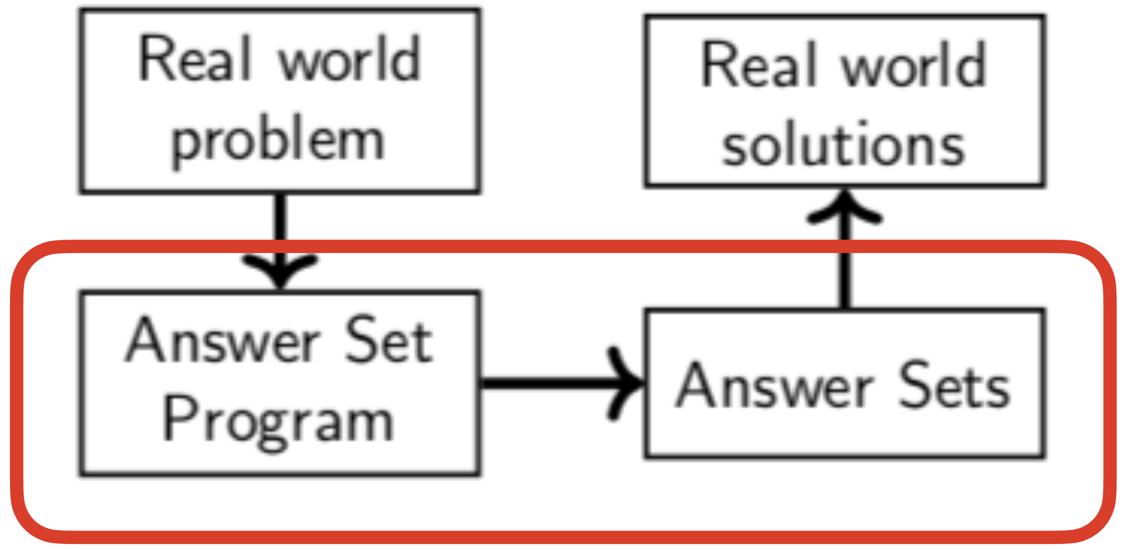
Imperial College London

Inductive Learning of Answer Set Programs

The ILASP Systems

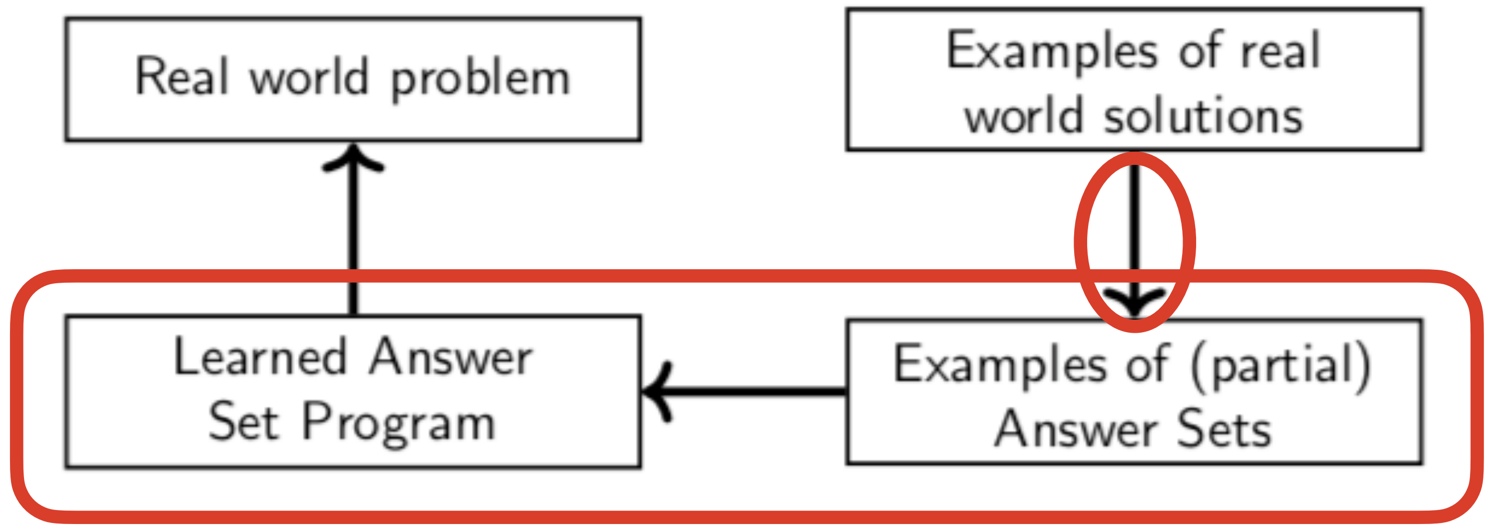
Answer Set Programming

Expressive Declarative Environment
for Reasoning Logically



ILASP

Expressive Declarative Environment
for Learning Logically



ILASP
Learning Logically

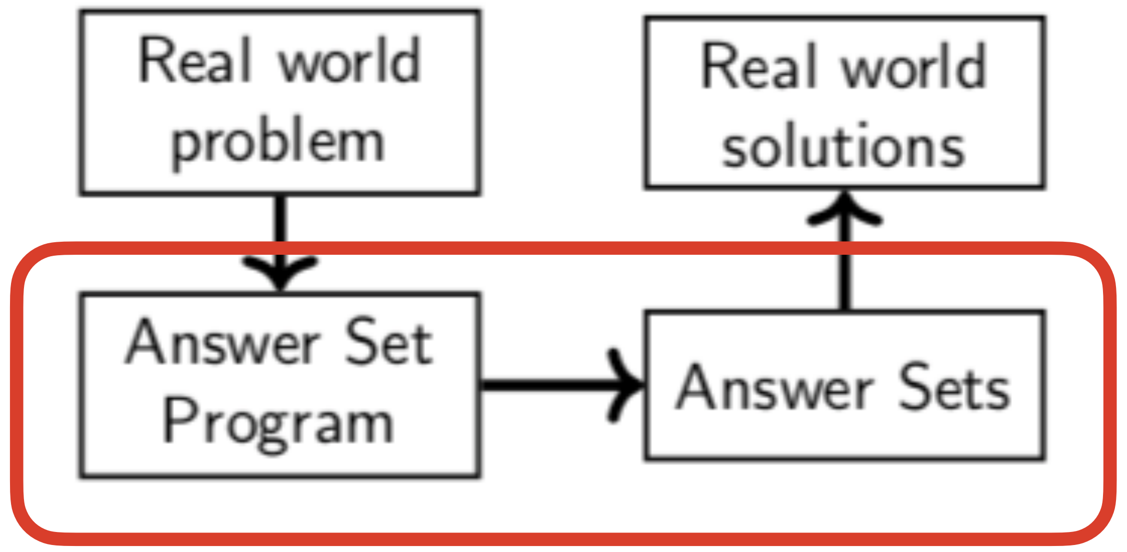
Imperial College London

Inductive Learning of Answer Set Programs

The ILASP Systems

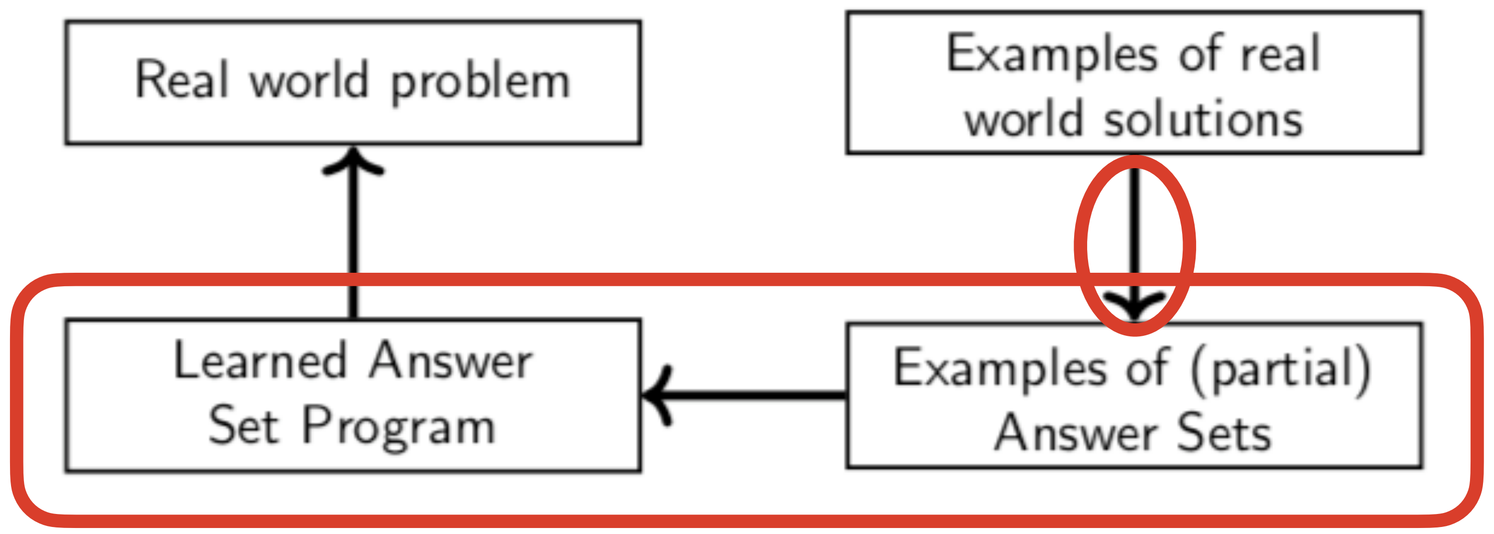
Answer Set Programming

Expressive Declarative Environment
for Reasoning Logically



ILASP

Expressive Declarative Environment
for Learning Logically



Desirable features for expressing (learned) knowledge in cognitive systems:

- Defaults and exceptions can be modelled using negation as failure
- Non-determinism and choice can modelled using choice rules
- Preferences can be modelled using weak constraints

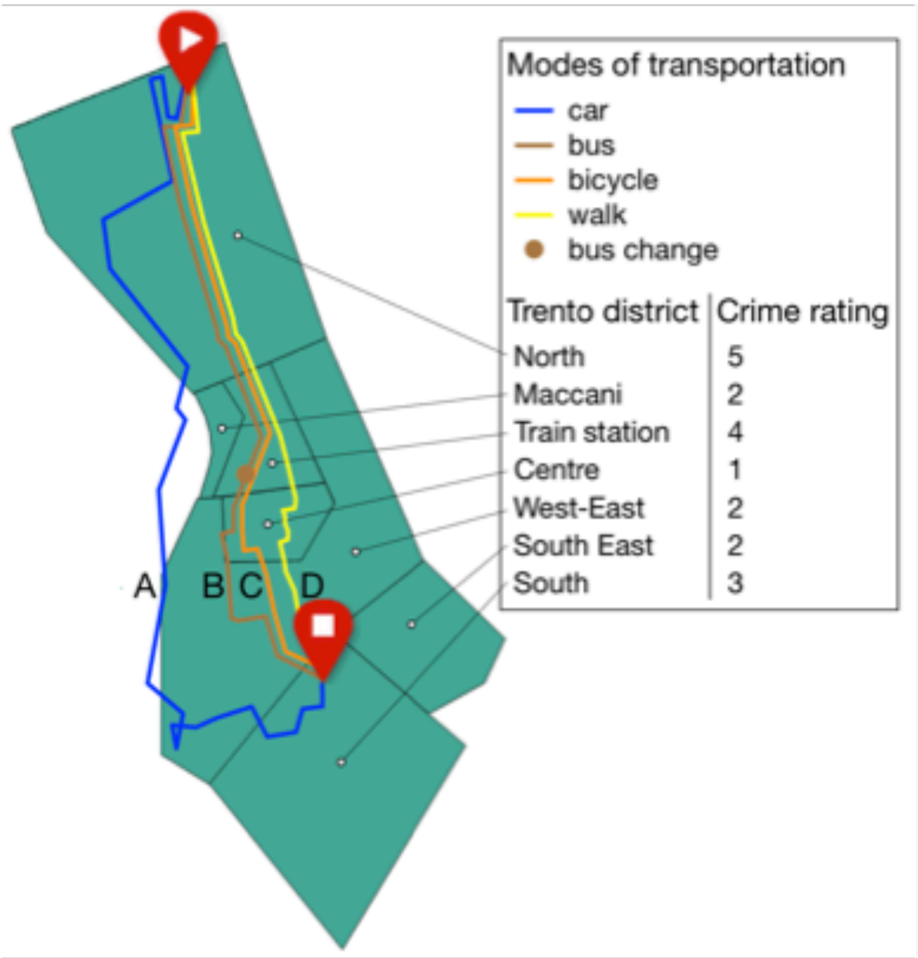
ILASP
Learning Logically

Learning Preference Models

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

Intelligent Urban Mobility System

ILASP2



ILASP
Learning Logically

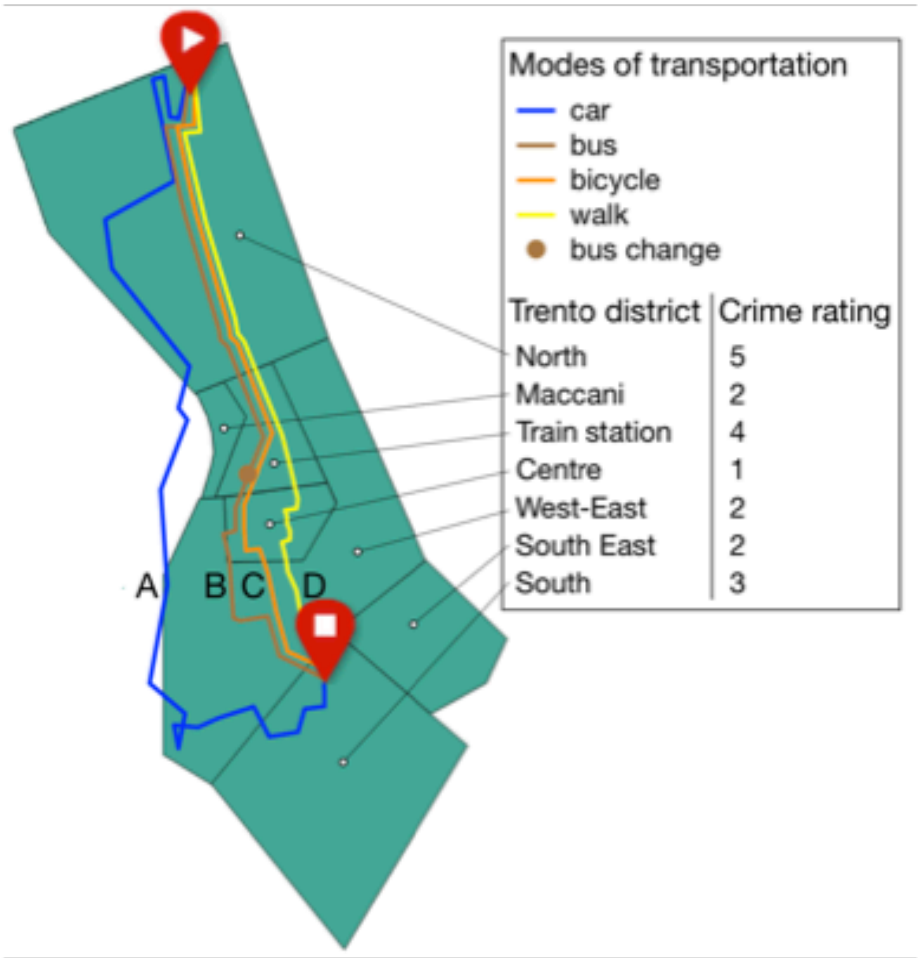
M. Law, A. Russo, K/ Broda. Learning weak constraints in answer set programming. TPLP 15(4-5), 2015.

Imperial College London

Learning Preference Models

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

Intelligent Urban Mobility System



ILASP2

Suggest user different alternatives:

Journey A	Journey B	Journey C	Journey D
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.	• Take the bus 400m through an area with crime rating of 2. • Take a second bus 3km through an area with crime rating 4	• Take a bus 2km through an area with crime rating 5. • Walk 2km through an area with crime rating 1.

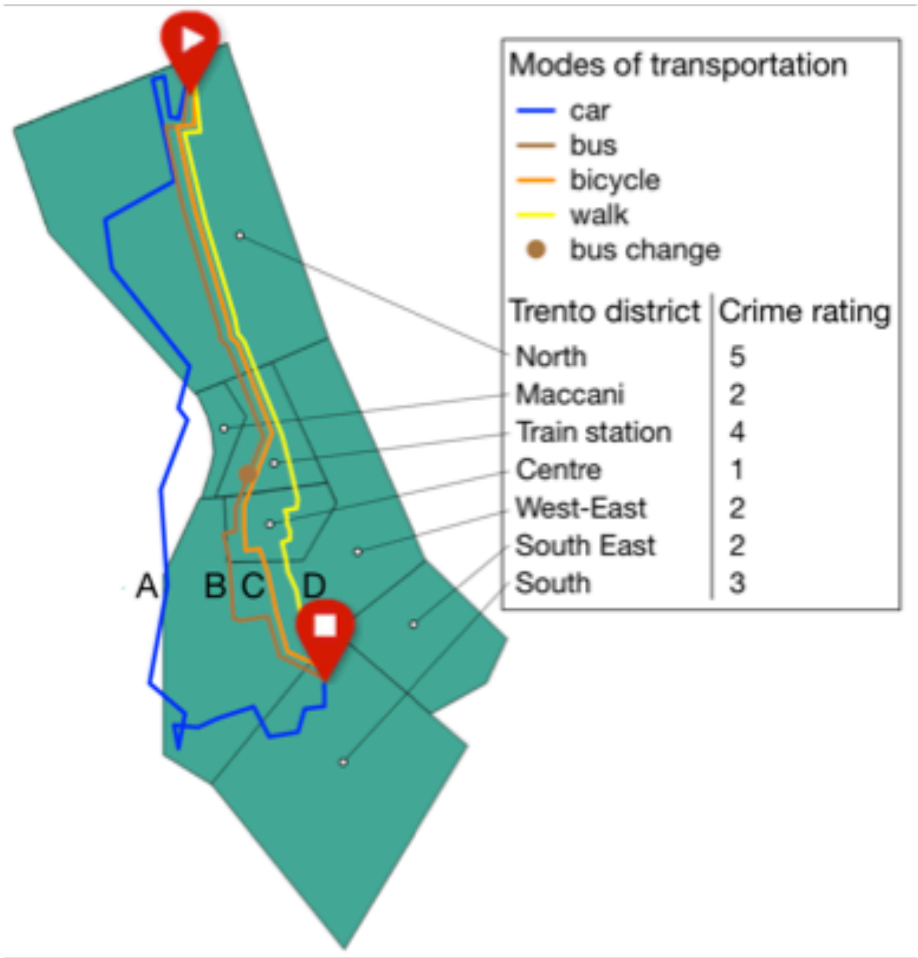
ILASP
Learning Logically

M. Law, A. Russo, K/ Broda. Learning weak constraints in answer set programming. TPLP 15(4-5), 2015.

Learning Preference Models

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

Intelligent Urban Mobility System



ILASP2

Suggest user different alternatives:

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

Journey C
- Take the bus 400m through an area with crime rating of 2.
- Take a second bus 3km through an area with crime rating 4

Journey D
- Take a bus 2km through an area with crime rating 5.
- Walk 2km through an area with crime rating 1.

User chooses A

Generate counter-examples

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

>

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

ILASP
Learning Logically

M. Law, A. Russo, K/ Broda. Learning weak constraints in answer set programming. TPLP 15(4-5), 2015.

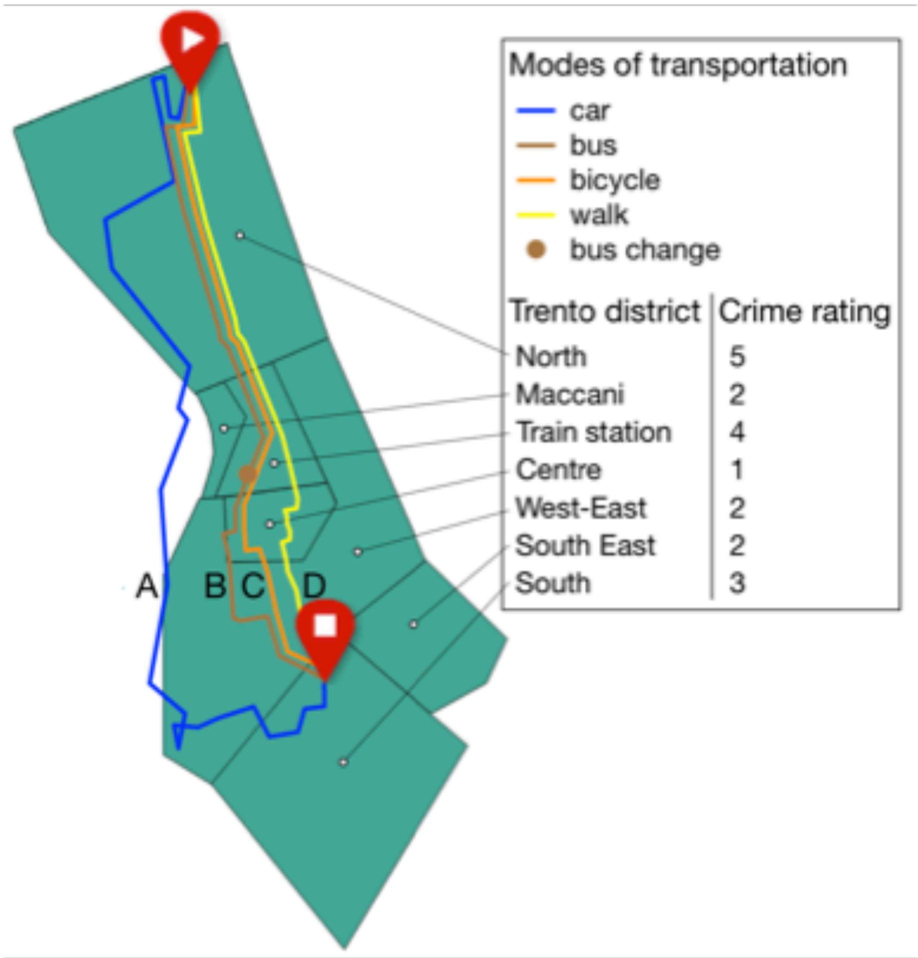
Imperial College London

Learning Preference Models

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

Intelligent Urban Mobility System

ILASP2



Suggest user different alternatives:

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

Journey C
- Take the bus 400m through an area with crime rating of 2.
- Take a second bus 3km through an area with crime rating 4

Journey D
- Take a bus 2km through an area with crime rating 5.
- Walk 2km through an area with crime rating 1.

User chooses A

Generate counter-examples

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

>

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

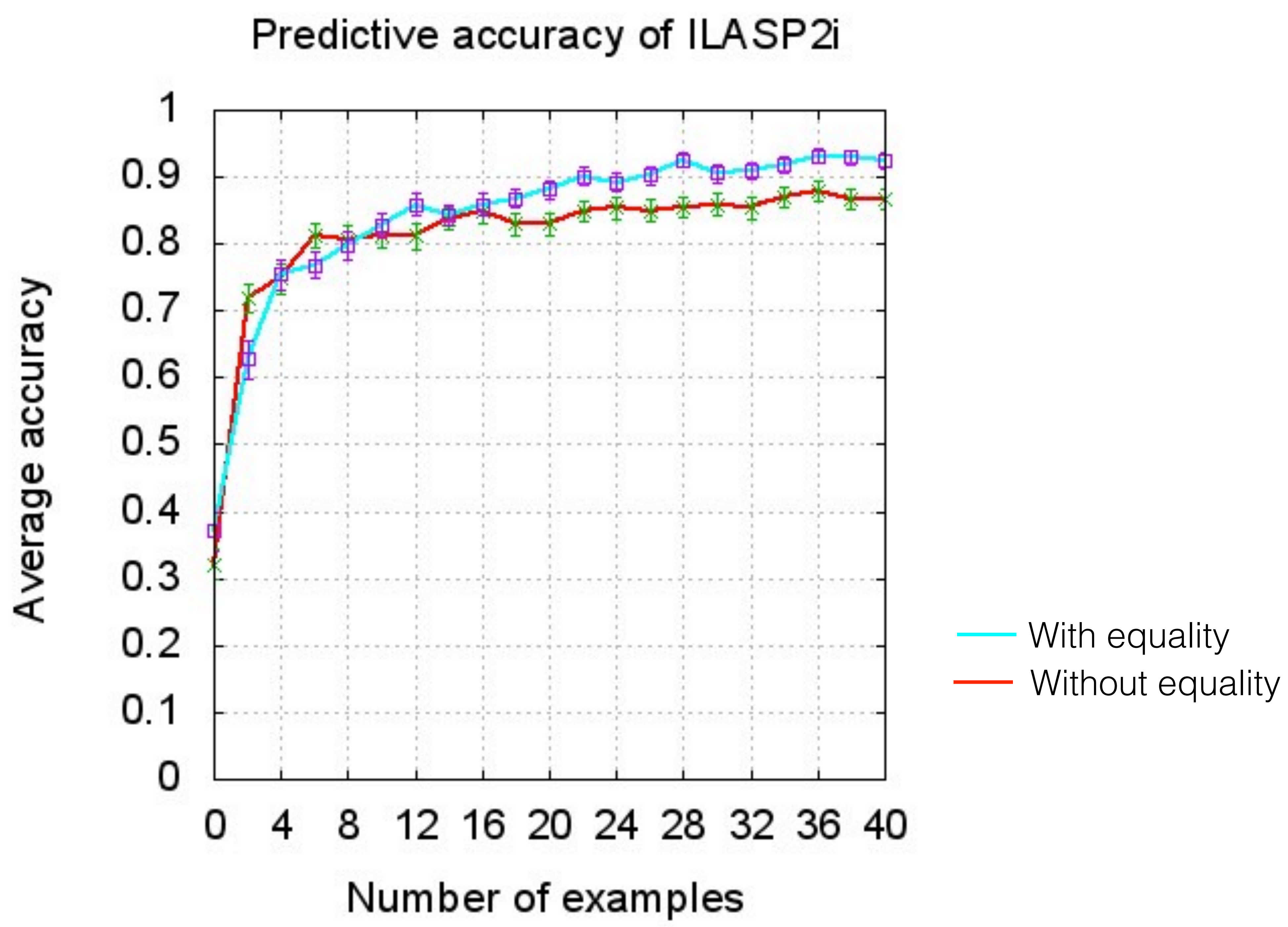
ILASP
Learning Logically

M. Law, A. Russo, K/ Broda. Learning weak constraints in answer set programming. TPLP 15(4-5), 2015.

Imperial College London

Learning Preference Models

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.



ILASP2

Suggest user different alternatives:

Journey A	Journey B	Journey C	Journey D
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.	• Take the bus 400m through an area with crime rating of 2. • Take a second bus 3km through an area with crime rating 4	• Take a bus 2km through an area with crime rating 5. • Walk 2km through an area with crime rating 1.

User chooses A

Generate counter-examples

Journey A	>	Journey B
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	>	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.

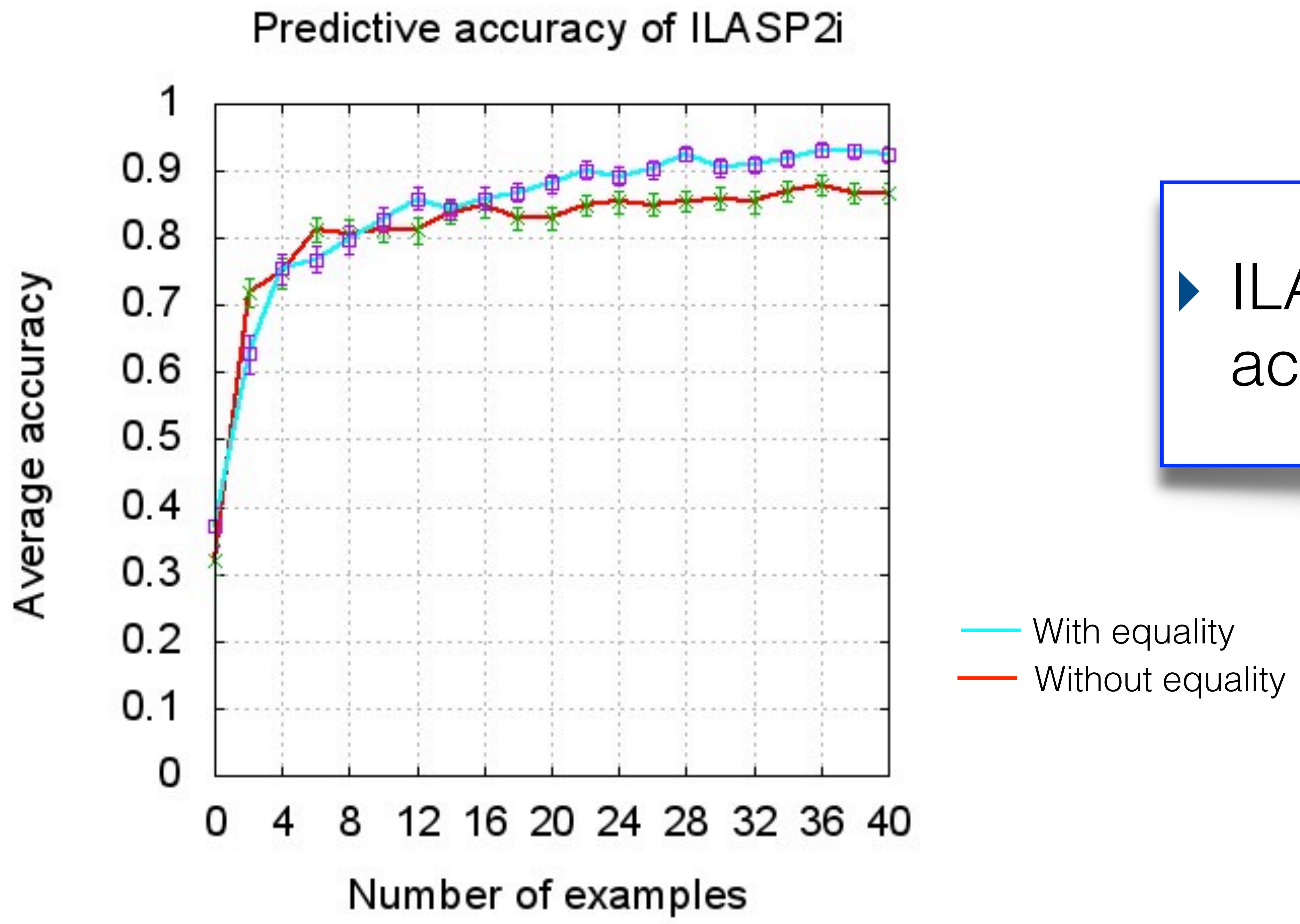
ILASP
Learning Logically

M. Law, A. Russo, K/ Broda. Learning weak constraints in answer set programming. TPLP 15(4-5), 2015.

Imperial College London

Learning Preference Models

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.



ILASP2

Suggest user different alternatives:

Journey A | Journey B | Journey C | Journey D

- ILASP2 is able to learn preferences, reaching 80% of accuracy from only 8 counter-examples.

User chooses A

Generate counter-examples

Journey A		Journey B
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	>	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.

ILASP
Learning Logically

M. Law, A. Russo, K/ Broda. Learning weak constraints in answer set programming. TPLP 15(4-5), 2015.

Learning from Context-dependent Examples

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

To scale up the number of counter-examples, context-dependent counter-examples can be considered.

ILASP2i

Suggest user different alternatives:

Journey A	Journey B	Journey C	Journey D
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.	• Take the bus 400m through an area with crime rating of 2. • Take a second bus 3km through an area with crime rating 4	• Take a bus 2km through an area with crime rating 5. • Walk 2km through an area with crime rating 1.

User chooses A

Generate counter-examples

Journey A		Journey B
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	>	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.

M. Law, A. Russo, K. Broda. Iterative Learning of Answer Set Programs from Context Dependent Examples. TPLP. 16(5-6), 2016.

ILASP Learning Logically

Learning from Context-dependent Examples

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

To scale up the number of counter-examples, context-dependent counter-examples can be considered.

ILASP2i

Suggest user different alternatives:

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

Journey C
- Take the bus 400m through an area with crime rating of 2.
- Take a second bus 3km through an area with crime rating 4

Journey D
- Take a bus 2km through an area with crime rating 5.
- Walk 2km through an area with crime rating 1.

User chooses A

Generate counter-examples

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

\>

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

if it is not raining

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

\>

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

if it is raining

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

<

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

M. Law, A. Russo, K. Broda. Iterative Learning of Answer Set Programs from Context Dependent Examples. TPLP. 16(5-6), 2016.

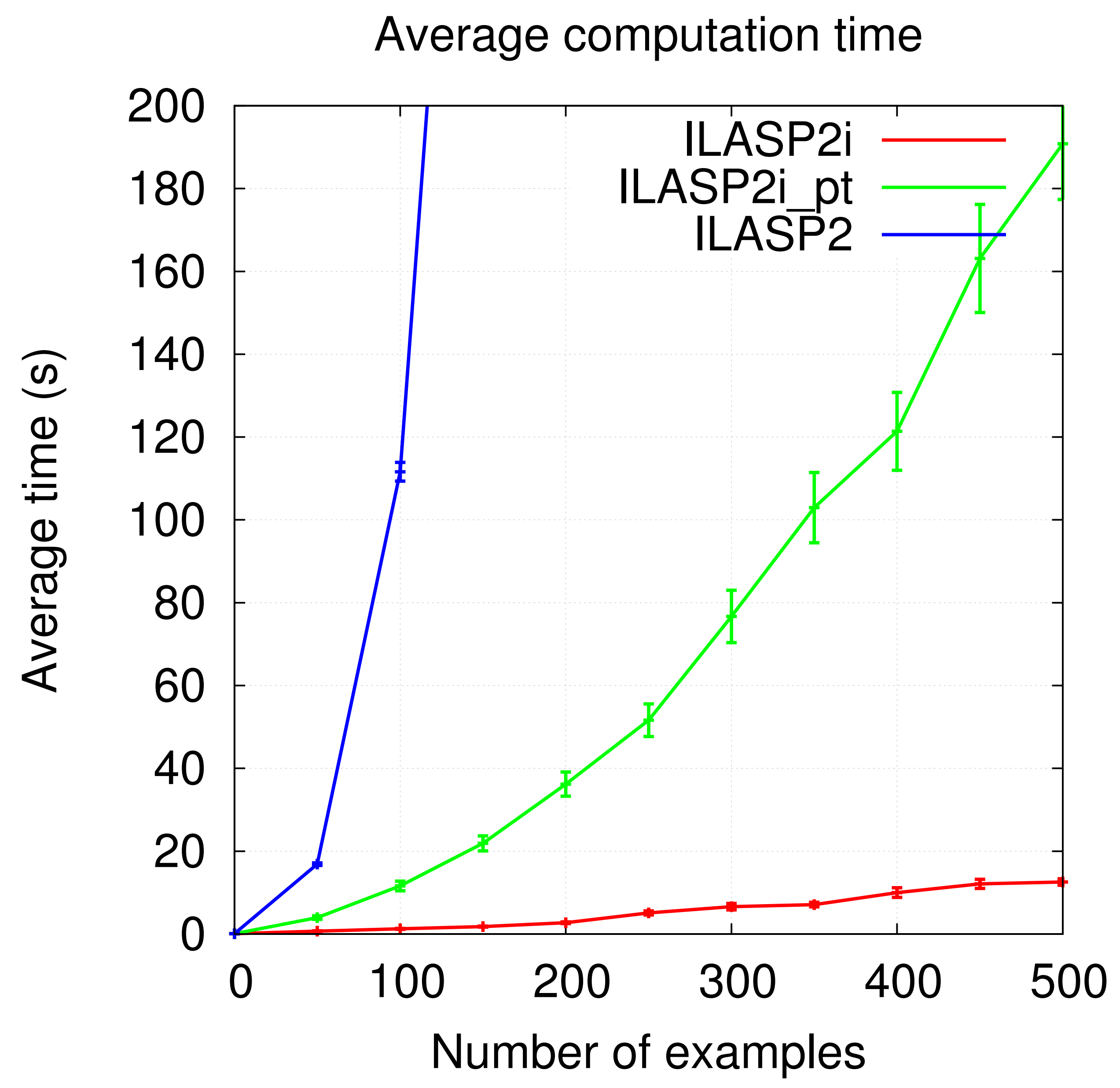
ILASP Learning Logically

Imperial College London

Learning from Context-dependent Examples

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

To scale up the number of counter-examples, context-dependent counter-examples can be considered.



ILASP2i

Suggest user different alternatives:

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

Journey C
- Take the bus 400m through an area with crime rating of 2.
- Take a second bus 3km through an area with crime rating 4

Journey D
- Take a bus 2km through an area with crime rating 5.
- Walk 2km through an area with crime rating 1.

User chooses A

Generate counter-examples

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

>

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

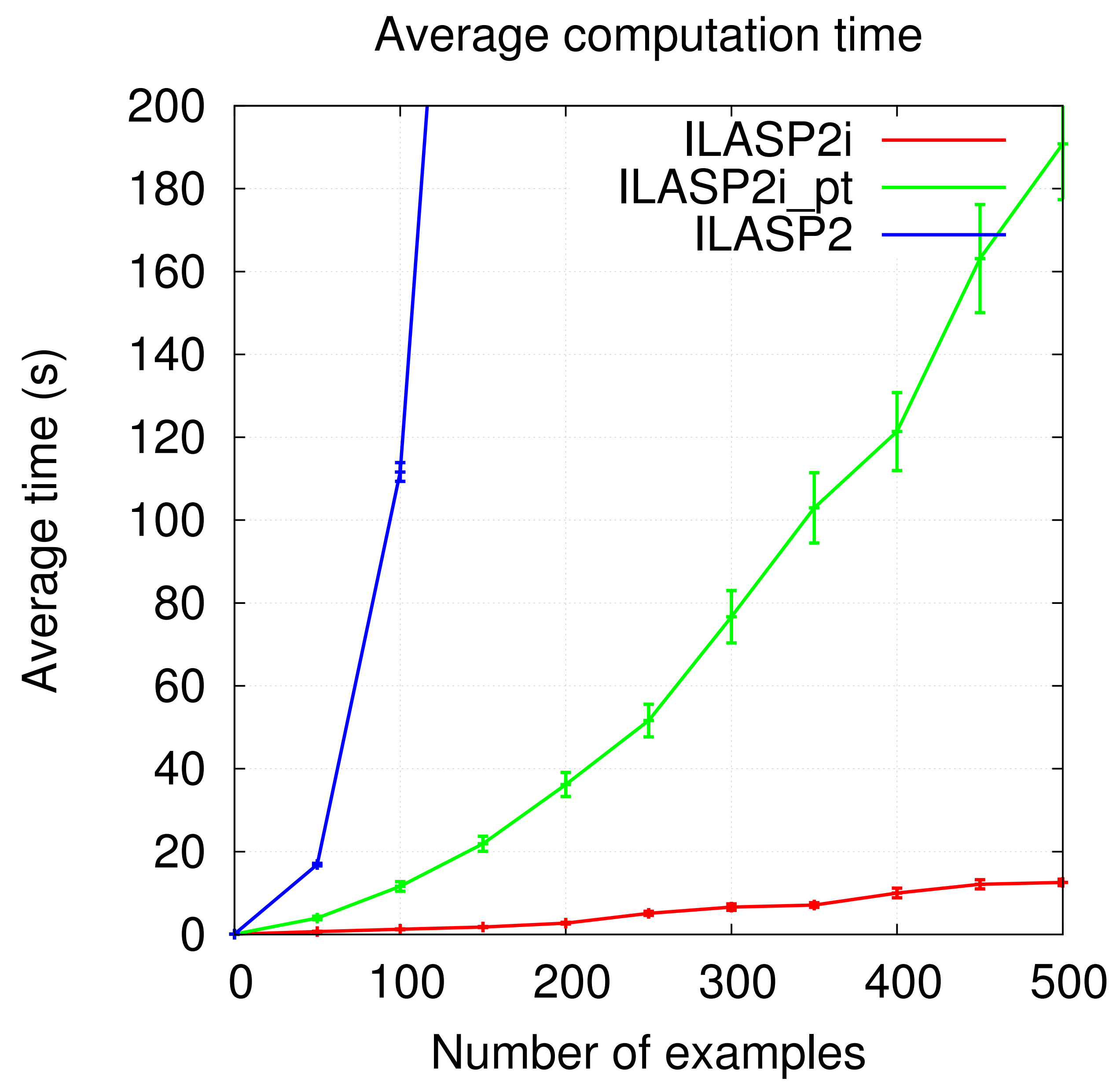
ILASP
Learning Logically

M. Law, A. Russo, K. Broda. Iterative Learning of Answer Set Programs from Context Dependent Examples. TPLP. 16(5-6), 2016.

Learning from Context-dependent Examples

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

To scale up the number of counter-examples, context-dependent counter-examples can be considered.



ILASP2i

Suggest user different alternatives:

Journey A | Journey B | Journey C | Journey D

- ILASP2i is able to learn preferences from large number of examples.

Generate counter-examples

User chooses A

Journey A		Journey B
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	>	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.

ILASP Learning Logically

M. Law, A. Russo, K. Broda. Iterative Learning of Answer Set Programs from Context Dependent Examples. TPLP. 16(5-6), 2016.

Imperial College
London

Learning from Noisy Examples

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

Counter-examples might be noisy as humans might not know what they prefer.

ILASP3

Suggest user different alternatives:

Journey A	Journey B	Journey C	Journey D
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.	• Take the bus 400m through an area with crime rating of 2. • Take a second bus 3km through an area with crime rating 4	• Take a bus 2km through an area with crime rating 5. • Walk 2km through an area with crime rating 1.

User chooses A

Generate counter-examples

Journey A		Journey B
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	>	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.

ILASP
Learning Logically

M. Law, A. Russo, K. Broda. Iterative Learning of Answer Set Programs from Noisy Examples. Advances in Cognitive Systems. Vol 6, 2018.

Imperial College London

Learning from Noisy Examples

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

Counter-examples might be noisy as humans might not know what they prefer.

mislabelled

if it is not raining

Journey A		Journey B
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	>	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.

if it is raining

Journey A		Journey B
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	<	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.

ILASP3

Suggest user different alternatives:

Journey A	Journey B	Journey C	Journey D
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.	• Take the bus 400m through an area with crime rating of 2. • Take a second bus 3km through an area with crime rating 4	• Take a bus 2km through an area with crime rating 5. • Walk 2km through an area with crime rating 1.

User chooses A

Generate counter-examples

Journey A		Journey B
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	>	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.

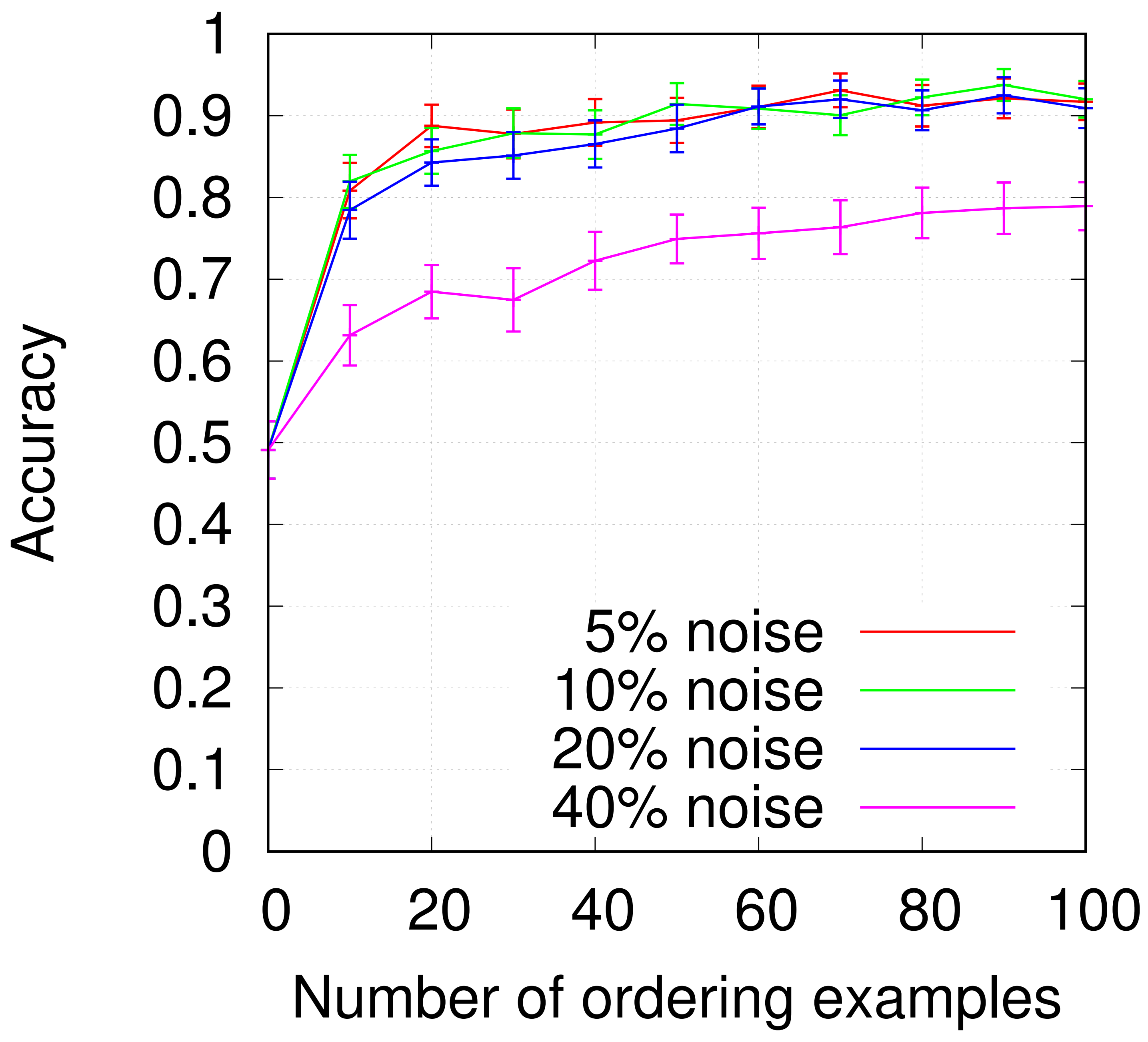
ILASP Learning Logically

M. Law, A. Russo, K. Broda. Iterative Learning of Answer Set Programs from Noisy Examples. Advances in Cognitive Systems. Vol 6, 2018.

Learning from Noisy Examples

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

Counter-examples might be noisy as humans might not know what they prefer.



(c)

ILASP3

Suggest user different alternatives:

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

Journey C
- Take the bus 400m through an area with crime rating of 2.
- Take a second bus 3km through an area with crime rating 4

Journey D
- Take a bus 2km through an area with crime rating 5.
- Walk 2km through an area with crime rating 1.

User chooses A

Generate counter-examples

Journey A
- Walk 2km through an area with crime rating of 2.
- Take the bus 3km through an area with crime rating 4.

>

Journey B
- Take the bus 4km through an area with crime rating of 2
- Walk 1km through an area with crime rating 5.

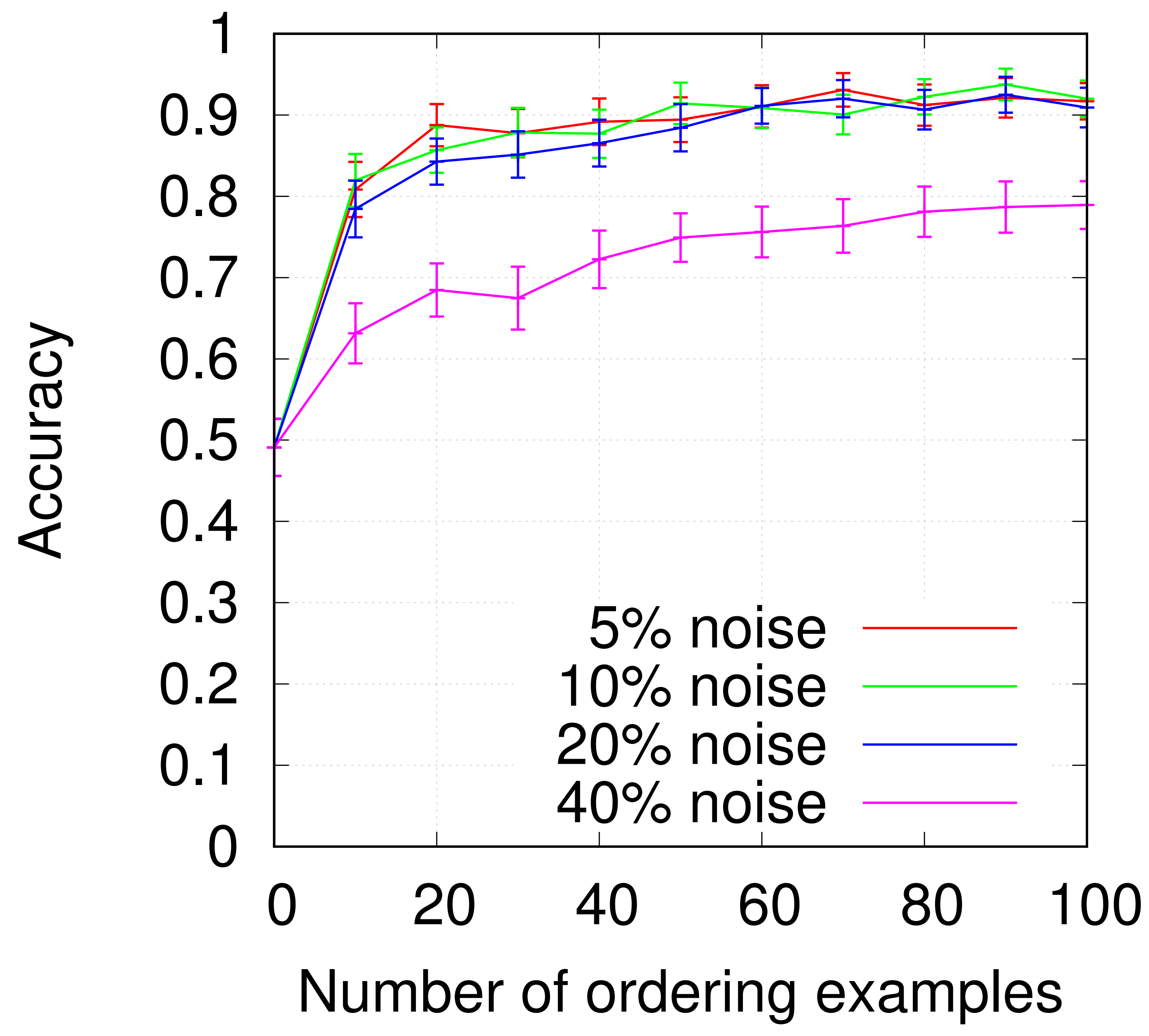
ILASP
Learning Logically

M. Law, A. Russo, K. Broda. Iterative Learning of Answer Set Programs from Noisy Examples. Advances in Cognitive Systems. Vol 6, 2018.

Learning from Noisy Examples

Objective is to learn human preferences from human's choices, and provide them with optimal, personalised suggestions with explanation.

Counter-examples might be noisy as humans might not know what they prefer.



(c)

ILASP3

Suggest user different alternatives:

Journey A	Journey B	Journey C	Journey D

- ILASP3 is as effective as ILASP2i, but able to learn from noisy examples.

User chooses A

Generate counter-examples

Journey A		Journey B
• Walk 2km through an area with crime rating of 2. • Take the bus 3km through an area with crime rating 4.	>	• Take the bus 4km through an area with crime rating of 2 • Walk 1km through an area with crime rating 5.

ILASP
Learning Logically

M. Law, A. Russo, K. Broda. Iterative Learning of Answer Set Programs from Noisy Examples. Advances in Cognitive Systems. Vol 6, 2018.

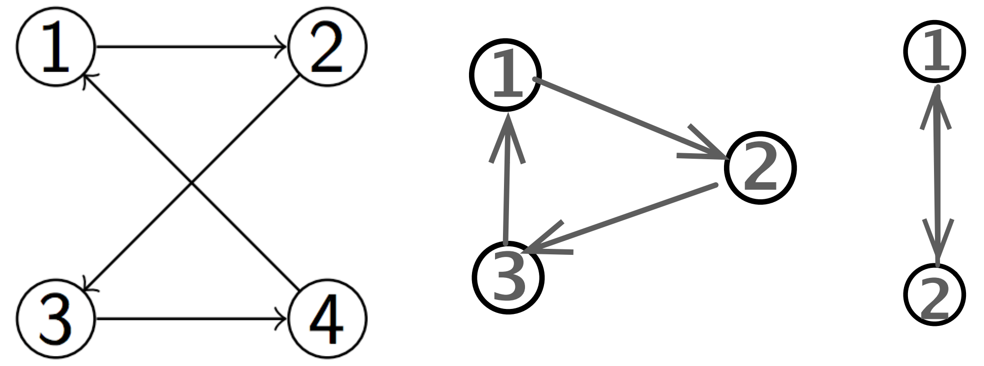
Learning Complex Definitions

Symbolic Machine Learning is highly declarative, and capable of learning definitions of complex (NP-hard) decision problems.

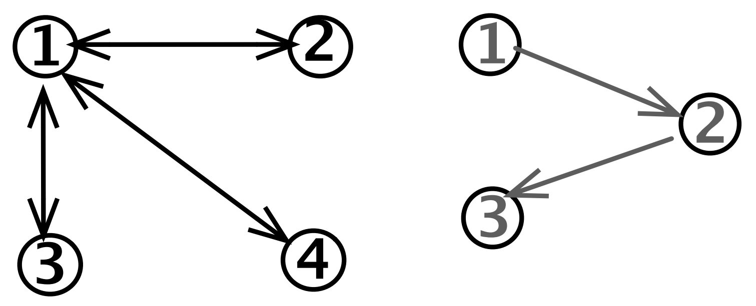
Background:

```
1{ size(1), size(2), size(3), size(4) }1.
node(1..S) :- size(S).
0{ edge(V0, V1) }1 :- node(V0), node(V1).
```

Positive examples:



Negative examples:



M. Law, A. Russo, K. Broda. The complexity and generality of learning answer set programs. Artificial Intelligence Journal, 2018.

ILASP
Learning Logically

Imperial College London

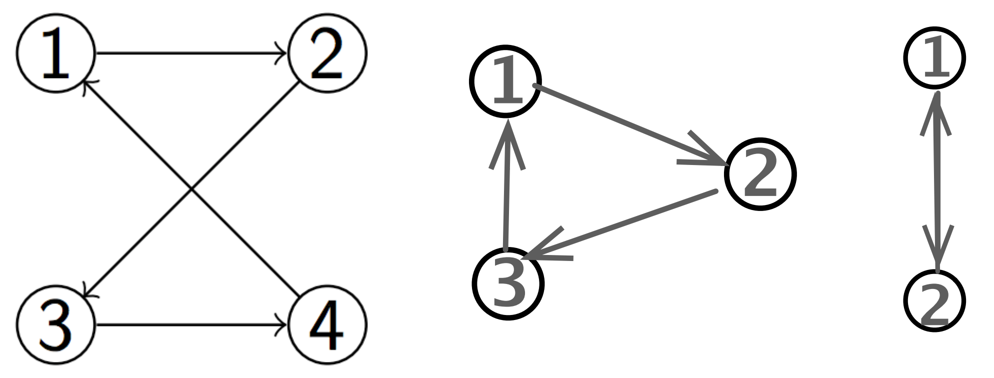
Learning Complex Definitions

Symbolic Machine Learning is highly declarative, and capable of learning definitions of complex (NP-hard) decision problems.

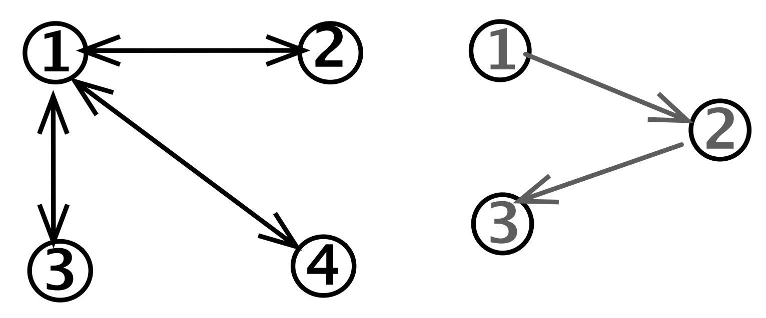
Background:

```
1{ size(1), size(2), size(3), size(4) }1.
node(1..S) :- size(S).
0{ edge(V0, V1) }1 :- node(V0), node(V1).
```

Positive examples:

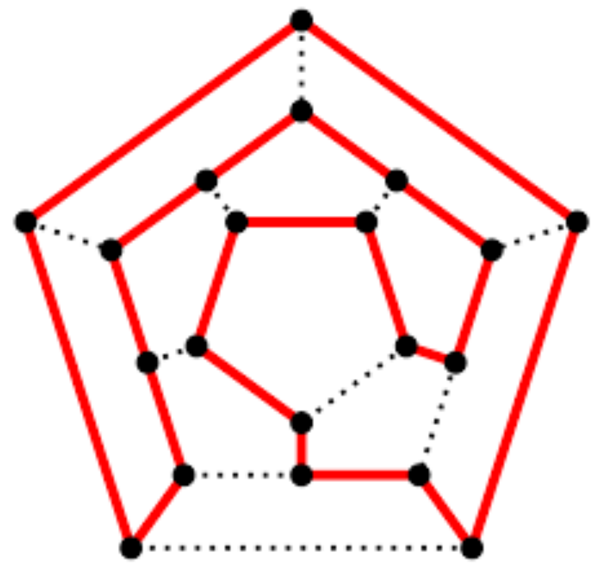


Negative examples:



Learned definition of Hamiltonian graph:

```
0{ in_hc(V0,V1) }1 :- edge(V0,V1)
reach(V0) :- in_hc(1,V0)
reach(V1) :- in_hc(V0,V1), reach(V0)
:- node(V0), not reach(V0)
:- in_hc(V0,V1), in_hc(V0,V2), V1≠ V2
```



ILASP
Learning Logically

M. Law, A. Russo, K. Broda. The complexity and generality of learning answer set programs. Artificial Intelligence Journal, 2018.

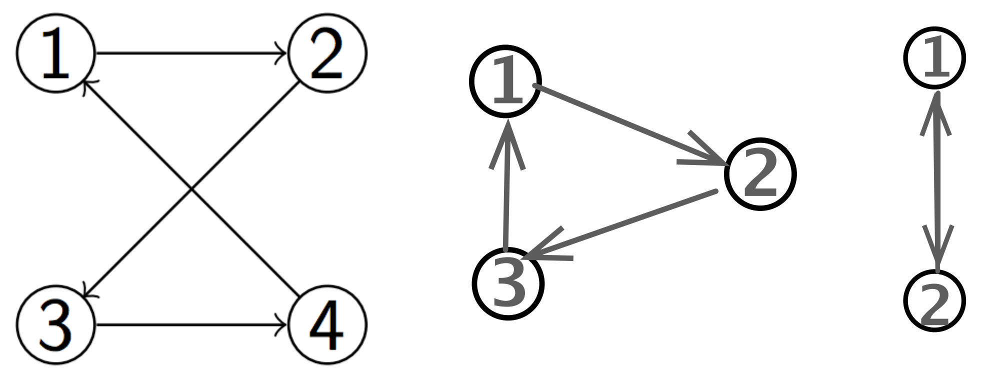
Learning Complex Definitions

Symbolic Machine Learning is highly declarative, and capable of learning definitions of complex (NP-hard) decision problems.

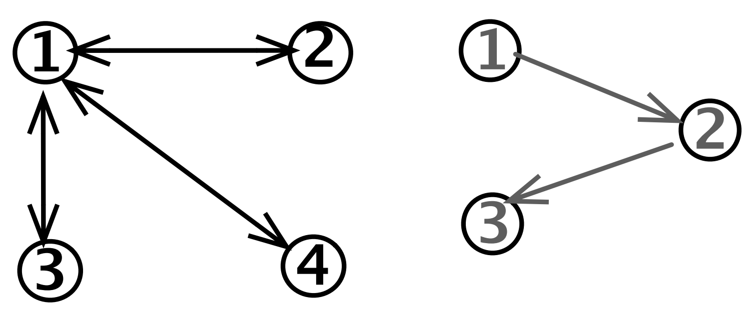
Background:

```
1{ size(1), size(2), size(3), size(4) }1.
node(1..S) :- size(S).
0{ edge(V0, V1) }1 :- node(V0), node(V1).
```

Positive examples:



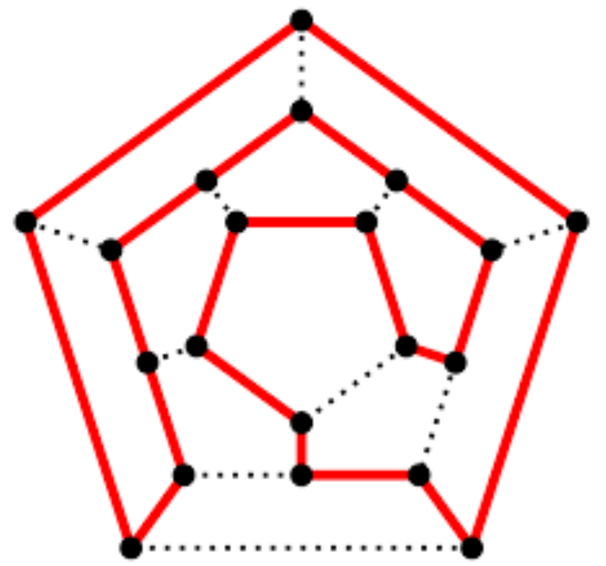
Negative examples:



Learned definition of Hamiltonian graph:

```
0{ in_hc(V0,V1) }1 :- edge(V0,V1)
reach(V0) :- in_hc(1,V0)
reach(V1) :- in_hc(V0,V1), reach(V0)
:- node(V0), not reach(V0)
:- in_hc(V0,V1), in_hc(V0,V2), V1≠ V2
```

A Hamilton cycle is a subset of the edges in the graph.



ILASP
Learning Logically

M. Law, A. Russo, K. Broda. The complexity and generality of learning answer set programs. Artificial Intelligence Journal, 2018.

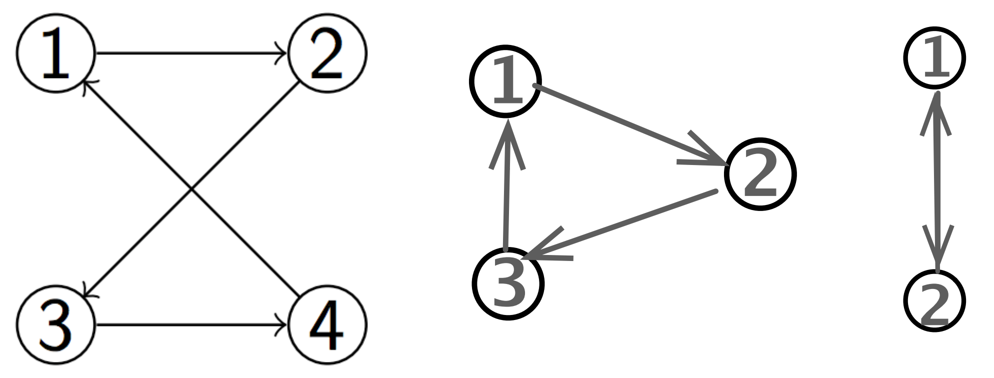
Learning Complex Definitions

Symbolic Machine Learning is highly declarative, and capable of learning definitions of complex (NP-hard) decision problems.

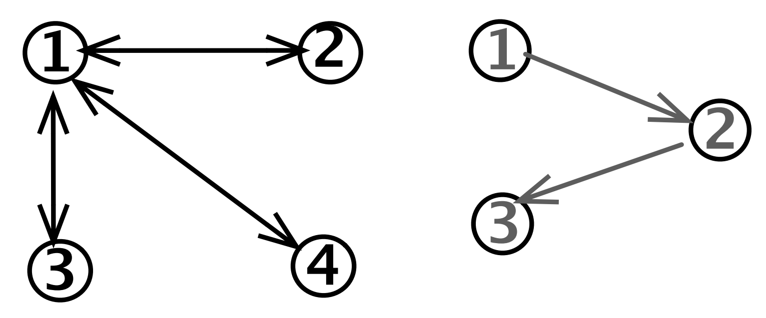
Background:

```
1{ size(1), size(2), size(3), size(4) }1.
node(1..S) :- size(S).
0{ edge(V0, V1) }1 :- node(V0), node(V1).
```

Positive examples:



Negative examples:

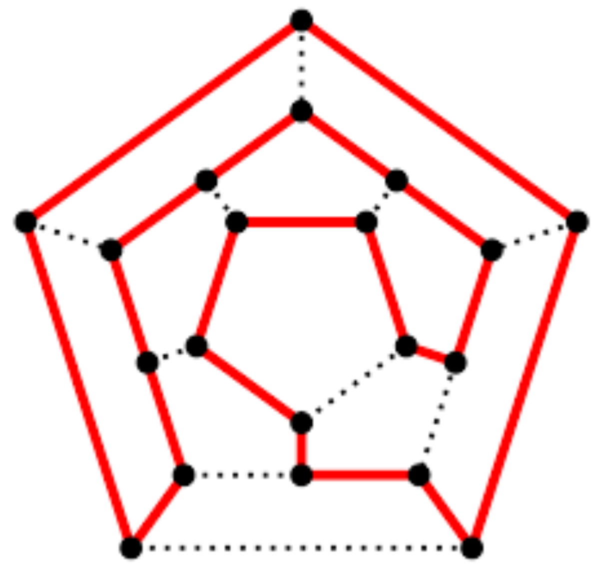


Learned definition of Hamiltonian graph:

```
0{ in_hc(V0,V1) }1 :- edge(V0,V1)
reach(V0) :- in_hc(1,V0)
reach(V1) :- in_hc(V0,V1), reach(V0)
:- node(V0), not reach(V0)
:- in_hc(V0,V1), in_hc(V0,V2), V1≠ V2
```

Node *n* is “reachable” if there is a path from node 1 to *n*.

A Hamilton cycle is a subset of the edges in the graph.



ILASP
Learning Logically

M. Law, A. Russo, K. Broda. The complexity and generality of learning answer set programs. Artificial Intelligence Journal, 2018.

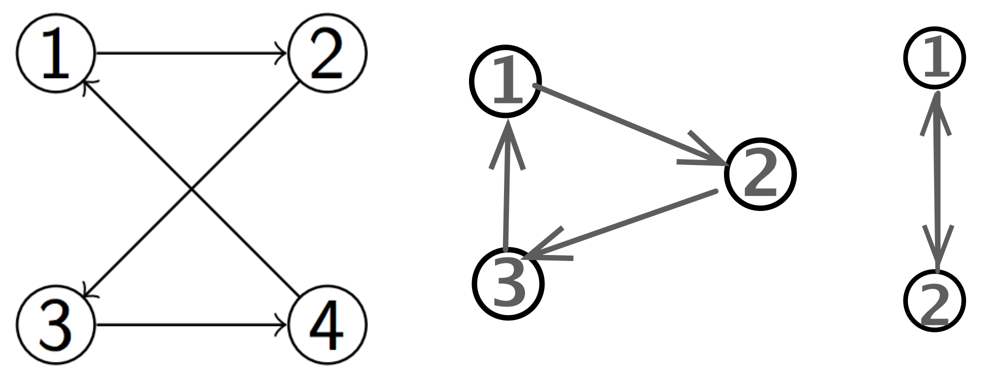
Learning Complex Definitions

Symbolic Machine Learning is highly declarative, and capable of learning definitions of complex (NP-hard) decision problems.

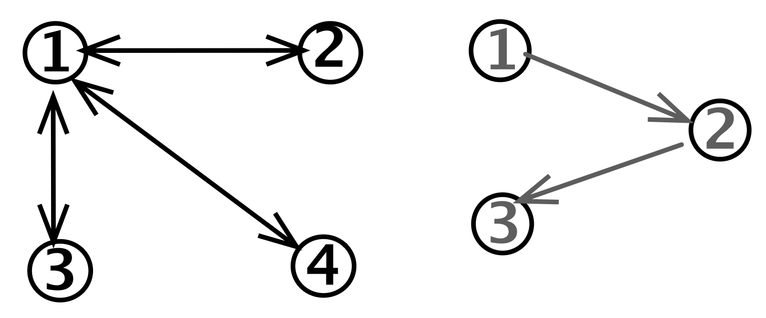
Background:

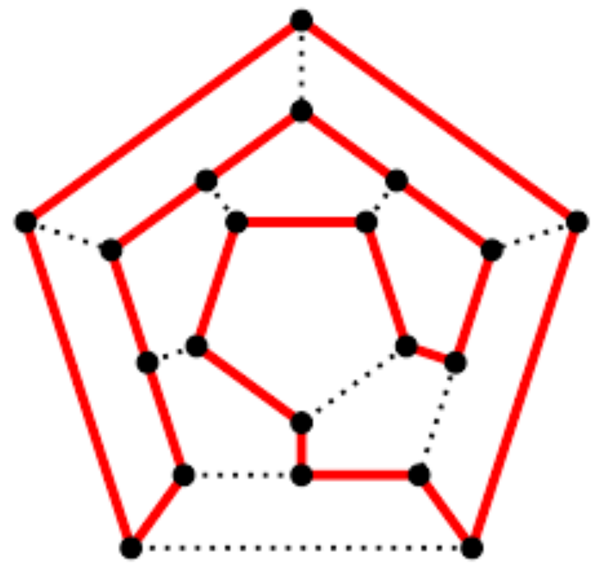
```
1{ size(1), size(2), size(3), size(4) }1.
node(1..S) :- size(S).
0{ edge(V0, V1) }1 :- node(V0), node(V1).
```

Positive examples:



Negative examples:





Learned definition of Hamiltonian graph:

```
0{ in_hc(V0,V1) }1 :- edge(V0,V1)
reach(V0) :- in_hc(1,V0)
reach(V1) :- in_hc(V0,V1), reach(V0)
:- node(V0), not reach(V0)
:- in_hc(V0,V1), in_hc(V0,V2), V1≠ V2
```

Node *n* is "reachable" if there is a path from node 1 to *n*.

A Hamilton cycle is a subset of the edges in the graph.

Every node must be reachable.

M. Law, A. Russo, K. Broda. The complexity and generality of learning answer set programs. Artificial Intelligence Journal, 2018.

ILASP Learning Logically

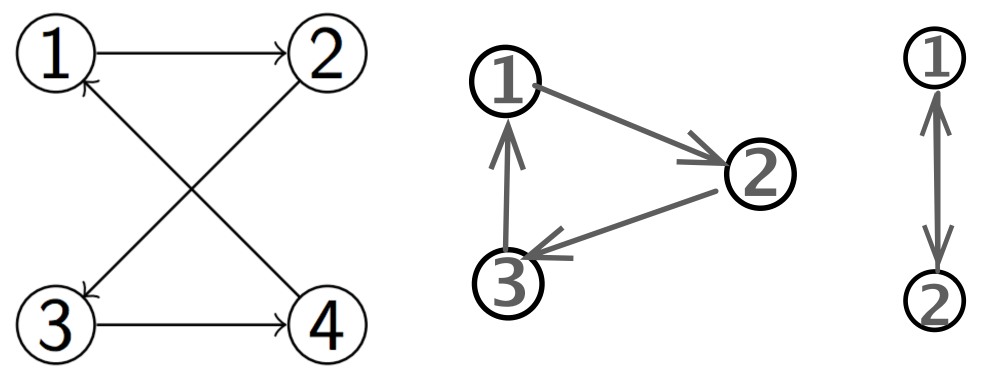
Learning Complex Definitions

Symbolic Machine Learning is highly declarative, and capable of learning definitions of complex (NP-hard) decision problems.

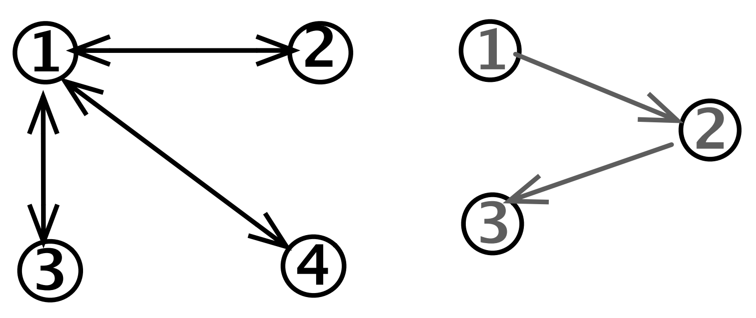
Background:

```
1{ size(1), size(2), size(3), size(4) }1.
node(1..S) :- size(S).
0{ edge(V0, V1) }1 :- node(V0), node(V1).
```

Positive examples:



Negative examples:



Learned definition of Hamiltonian graph:

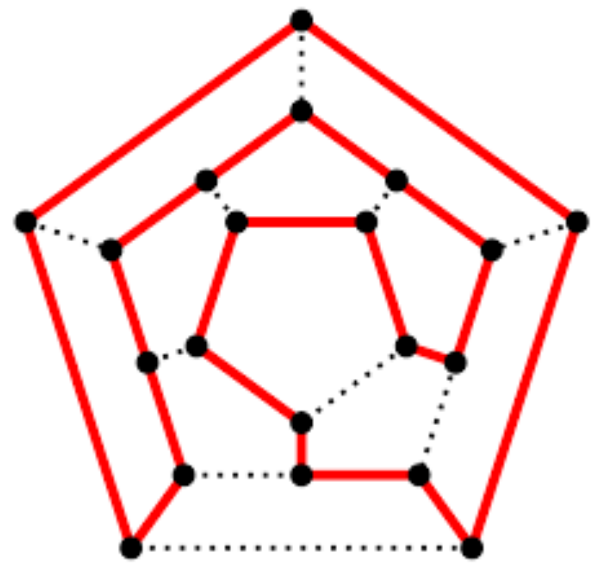
```
0{ in_hc(V0,V1) }1 :- edge(V0,V1)
reach(V0) :- in_hc(1,V0)
reach(V1) :- in_hc(V0,V1), reach(V0)
:- node(V0), not reach(V0)
:- in_hc(V0,V1), in_hc(V0,V2), V1≠ V2
```

Node *n* is “reachable” if there is a path from node 1 to *n*.

No node has more than one outgoing edge in the cycle.

A Hamilton cycle is a subset of the edges in the graph.

Every node must be reachable.

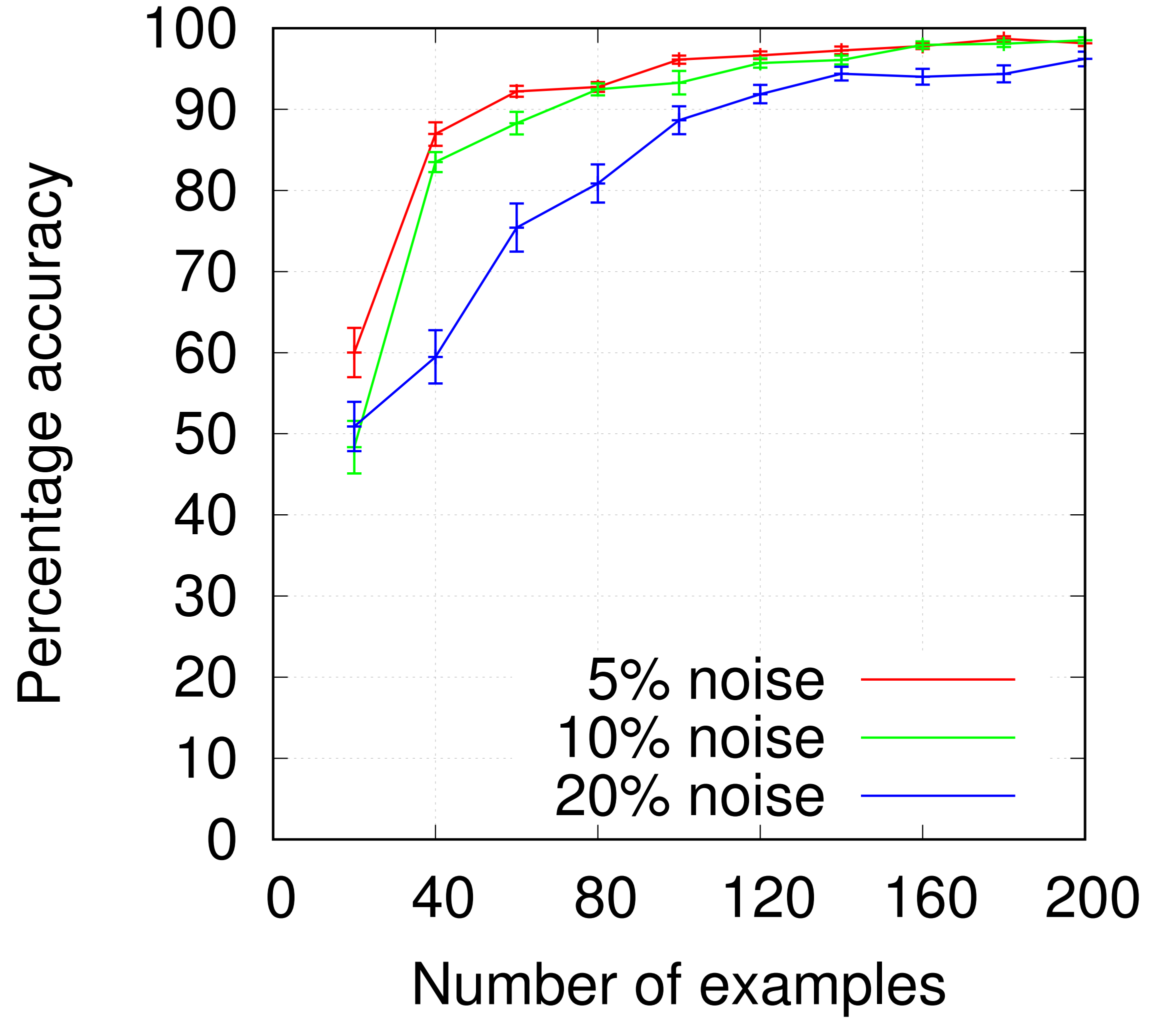


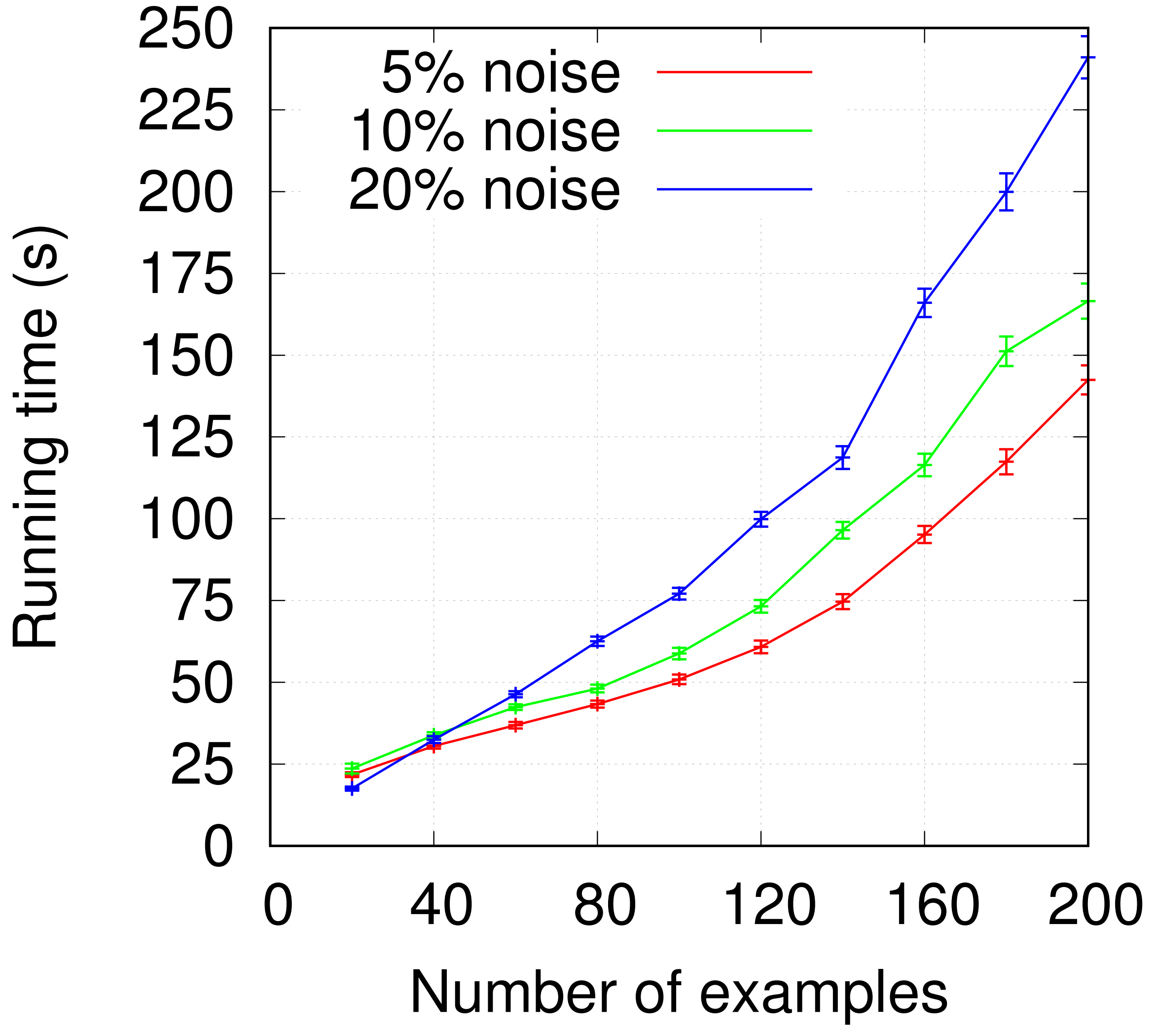
ILASP Learning Logically

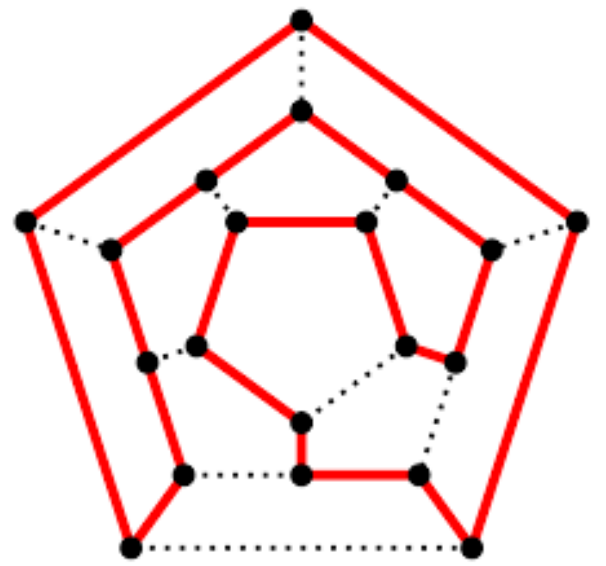
M. Law, A. Russo, K. Broda. The complexity and generality of learning answer set programs. Artificial Intelligence Journal, 2018.

Learning Complex Definitions

Symbolic Machine Learning is highly declarative, and capable of learning definitions of complex (NP-hard) decision problems.







Node *n* is "reachable" if there is a path from node 1 to *n*.

No node has more than one outgoing edge in the cycle.

```
0{ in_hc(V0,V1) }1 :- edge(V0,V1)
reach(V0) :- in_hc(1,V0)
reach(V1) :- in_hc(V0,V1), reach(V0)
:- node(V0), not reach(V0)
:- in_hc(V0,V1), in_hc(V0,V2), V1≠ V2
```

A Hamilton cycle is a subset of the edges in the graph.

Every node must be reachable.

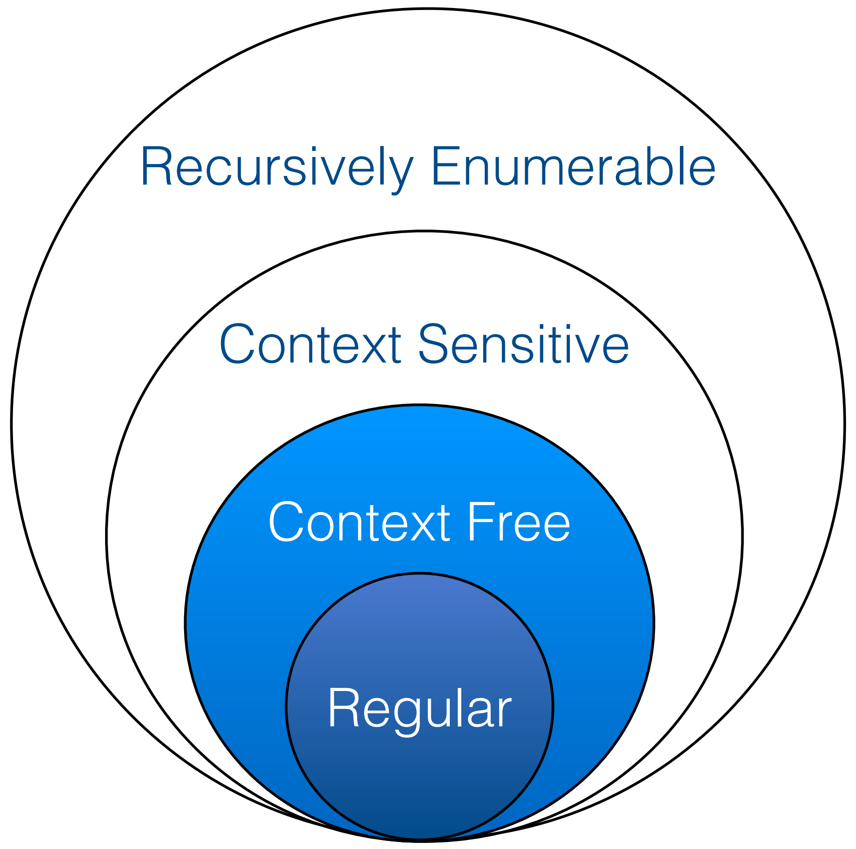
M. Law, A. Russo, K. Broda. The complexity and generality of learning answer set programs. Artificial Intelligence Journal, 2018.

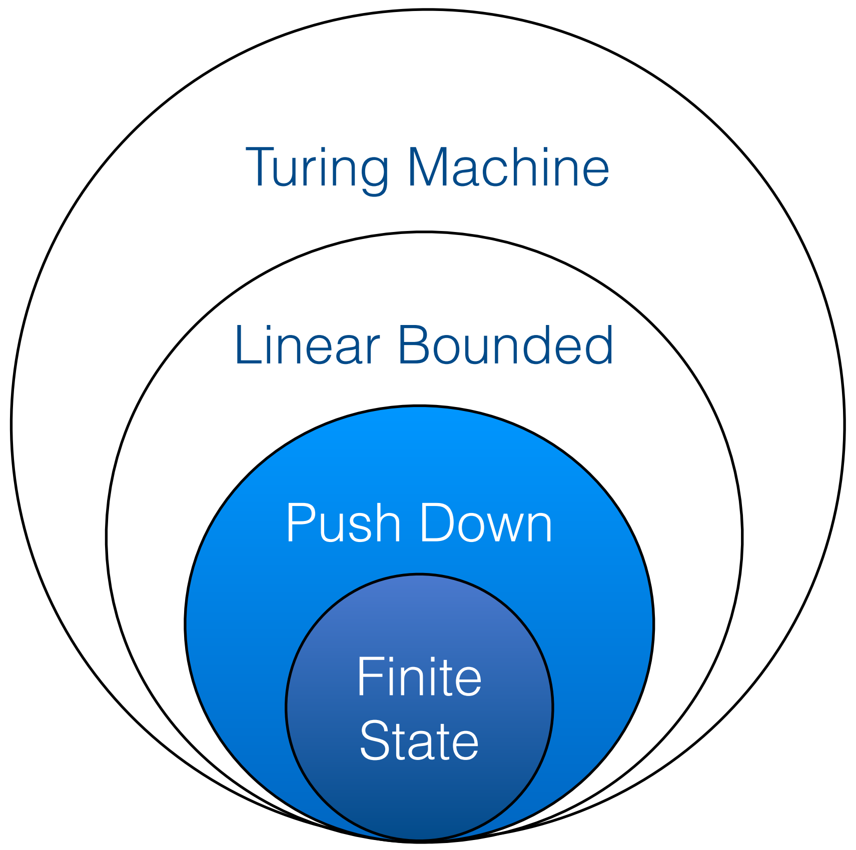
ILASP
Learning Logically

Imperial College London

Learning Complex Grammars

Previous work on learning grammars and automata has mostly been restricted to Regular Grammars (FSA) and Context-free Grammars (PDA).



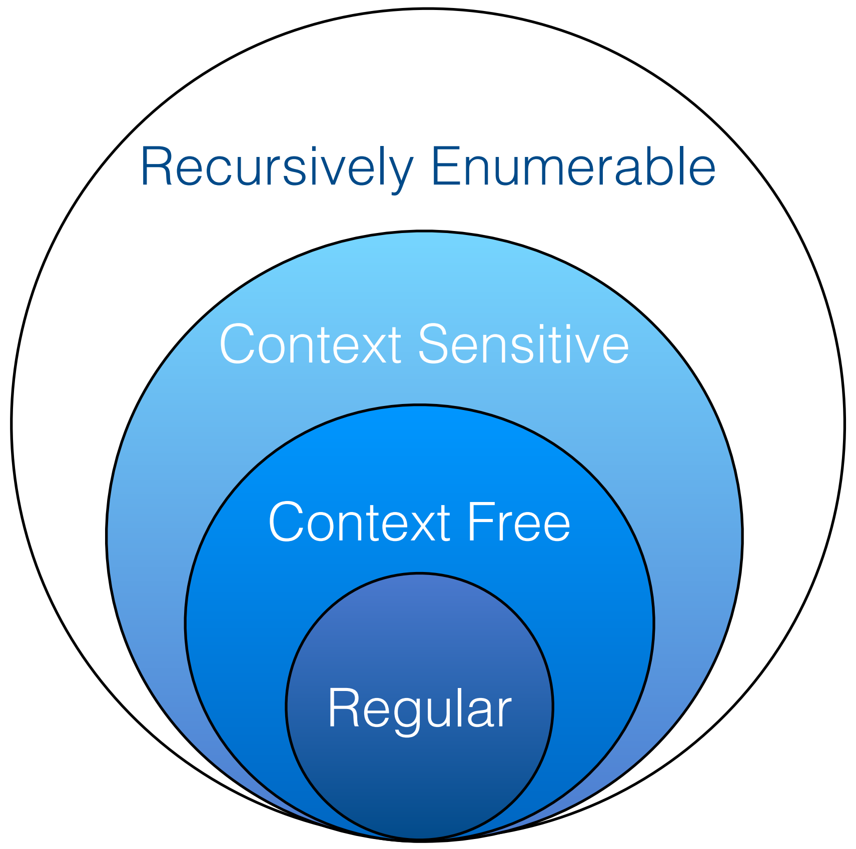


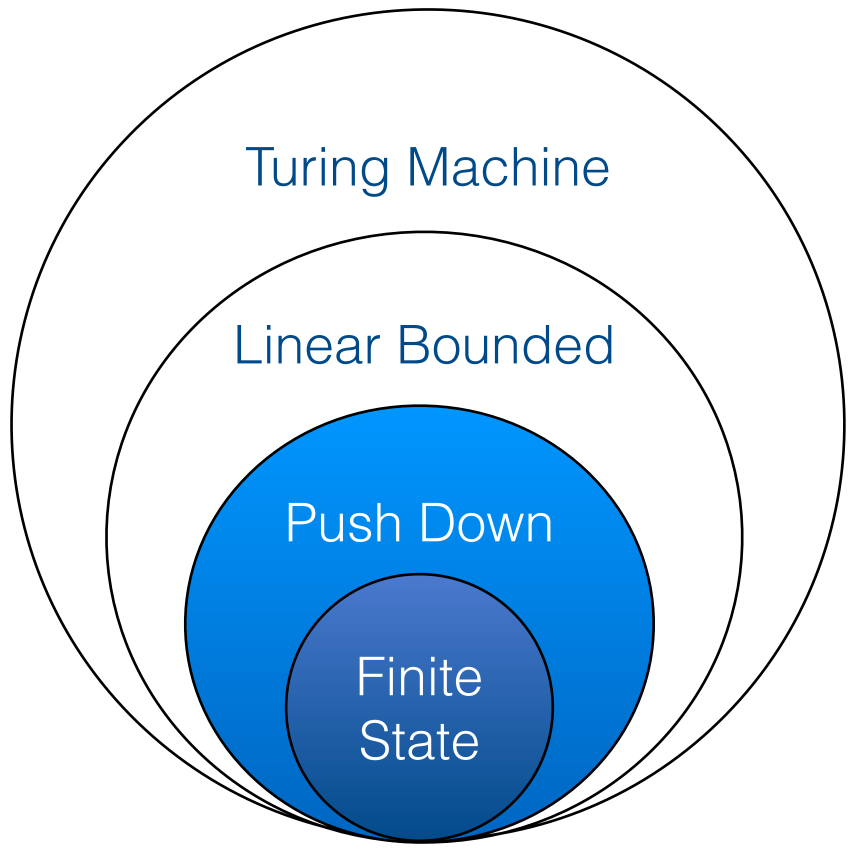
M. Law, A. Russo, E. Bertino, K.Broda, K. Lobo. Representing and Learning Grammars in Answer Set Programming, AAAI19.

Imperial College London

Learning Complex Grammars

Previous work on learning grammars and automata has mostly been restricted to Regular Grammars (FSA) and Context-free Grammars (PDA)





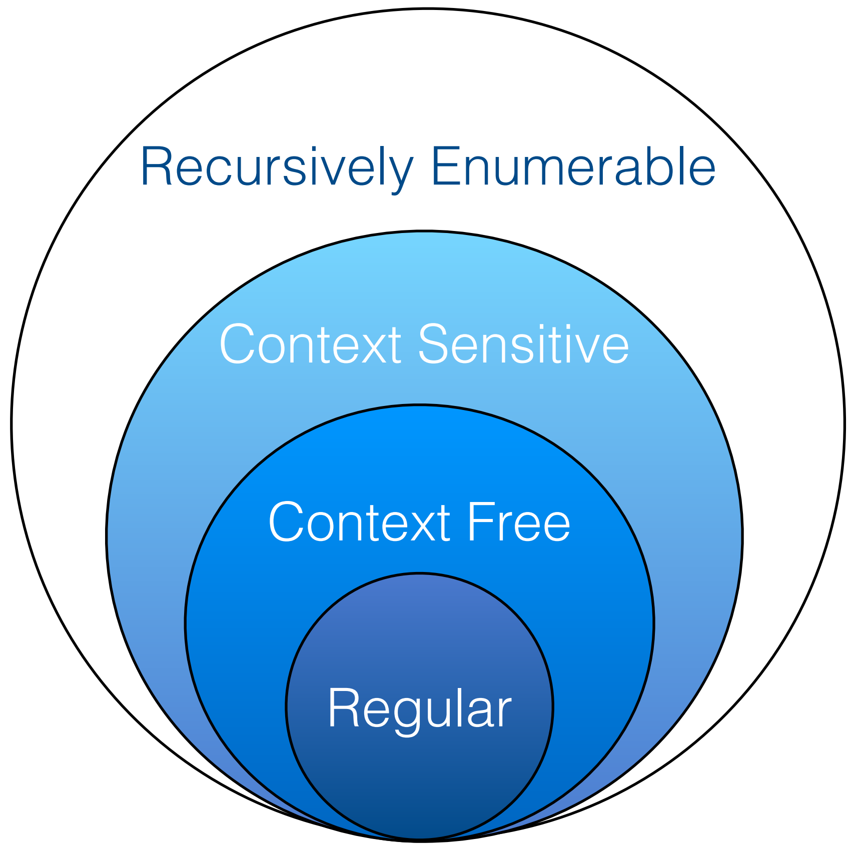
Learn a class of context-sensitive grammars (ASG):

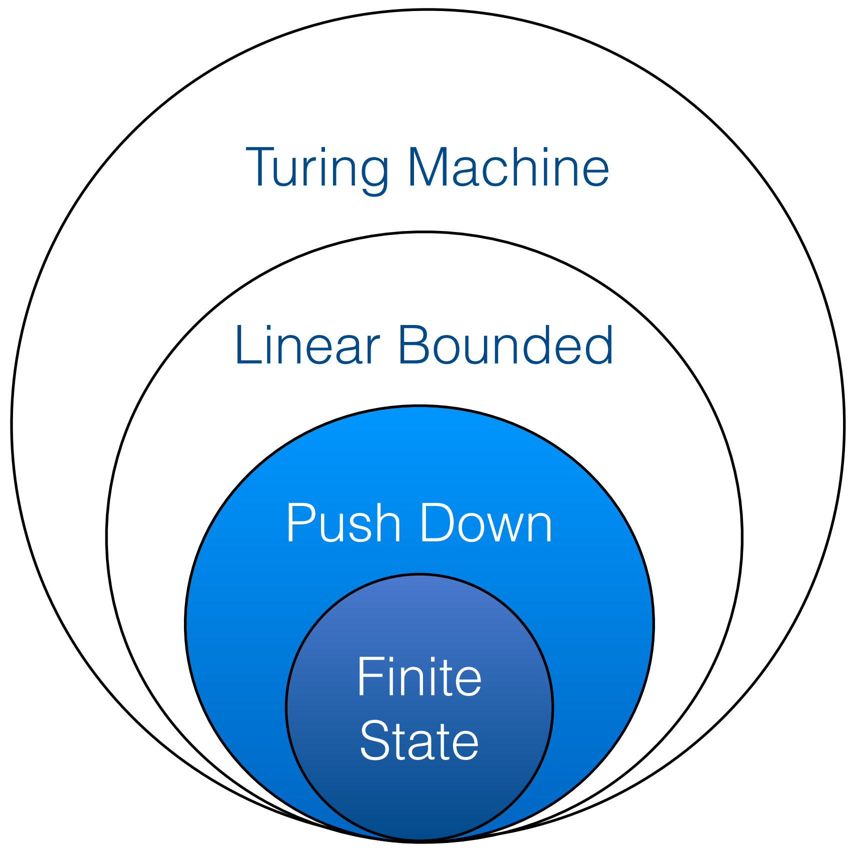
- context-free part defines the syntax of the language
- context-sensitive parts defines semantics.

M. Law, A. Russo, E. Bertino, K.Broda, K. Lobo. Representing and Learning Grammars in Answer Set Programming, AAAI19.

Learning Complex Grammars

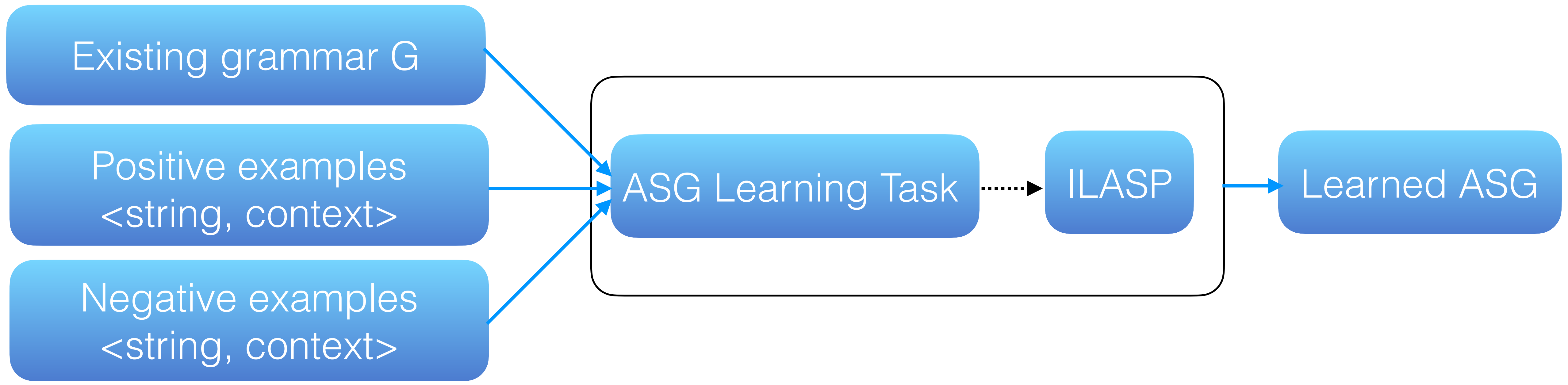
Previous work on learning grammars and automata has mostly been restricted to Regular Grammars (FSA) and Context-free Grammars (PDA)





Learn a class of context-sensitive grammars (ASG):

- context-free part defines the syntax of the language
- context-sensitive parts defines semantics.



Answer Set Grammars

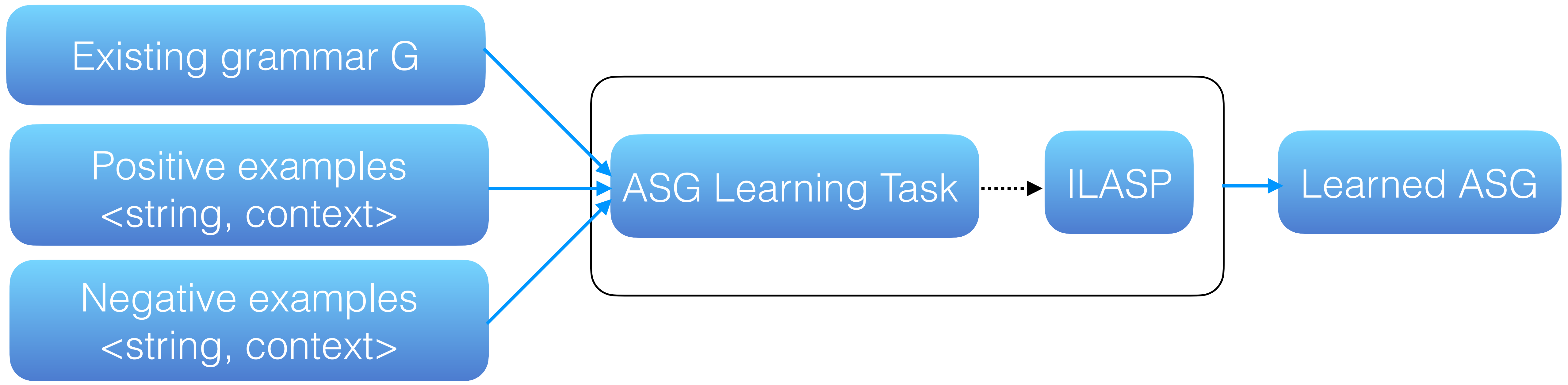
M. Law, A. Russo, E. Bertino, K.Broda, K. Lobo. Representing and Learning Grammars in Answer Set Programming, AAAI19.

Learning Complex Grammars

Example of context-sensitive grammar

Learn a class of context-sensitive grammars (ASG):

- context-free part defines the syntax of the language
- context-sensitive parts defines semantics.



Answer Set Grammars

M. Law, A. Russo, E. Bertino, K.Broda, K. Lobo. Representing and Learning Grammars in Answer Set Programming, AAAI19.

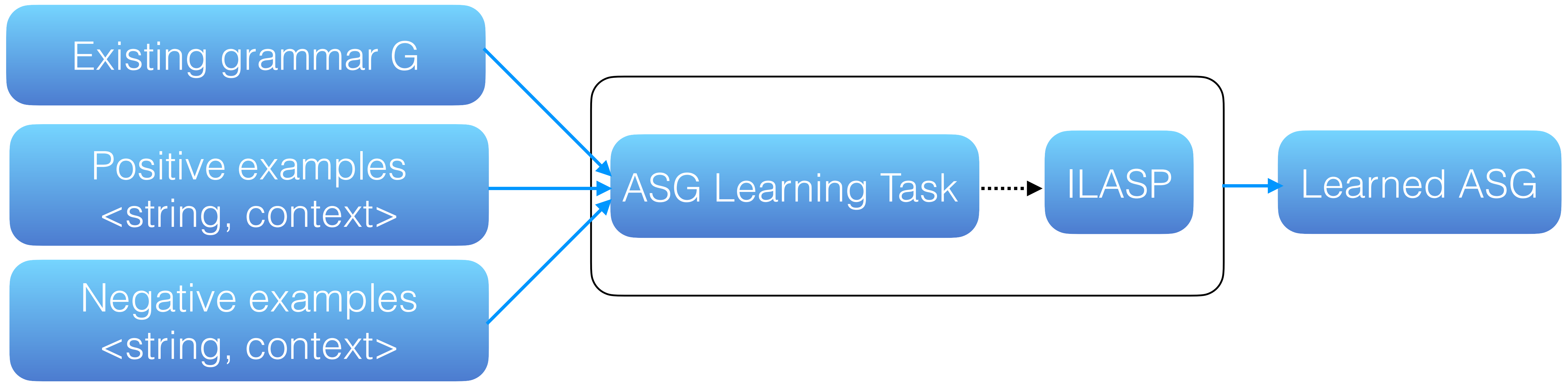
Learning Complex Grammars

Example of context-sensitive grammar

```
start -> as bs cs { false ← size(X)@1, not size(X)@2
                    false ← size(X)@1, not size(X)@2 }
as -> “a” as        { size(X+1) ← size(X)@2 }
as ->               { size(0) }
bs -> “b” bs        { size(X+1) ← size(X)@2 }
bs ->               { size(0) }
cs -> “c” cs        { size(X+1) ← size(X)@2 }
cs ->               { size(0) }
```

Learn a class of context-sensitive grammars (ASG):

- context-free part defines the syntax of the language
- context-sensitive parts defines semantics.



Answer Set Grammars

M. Law, A. Russo, E. Bertino, K.Broda, K. Lobo. Representing and Learning Grammars in Answer Set Programming, AAAI19.

Learning Complex Grammars

Example of context-sensitive grammar

```
start -> as bs cs {  false ← size(X)@1, not size(X)@2
                     false ← size(X)@1, not size(X)@2 }
as -> "a" as         { size(X+1) ← size(X)@2 }
as ->                { size(0) }
bs -> "b" bs         { size(X+1) ← size(X)@2 }
bs ->                { size(0) }
cs -> "c" cs         { size(X+1) ← size(X)@2 }
cs ->                { size(0) }
```

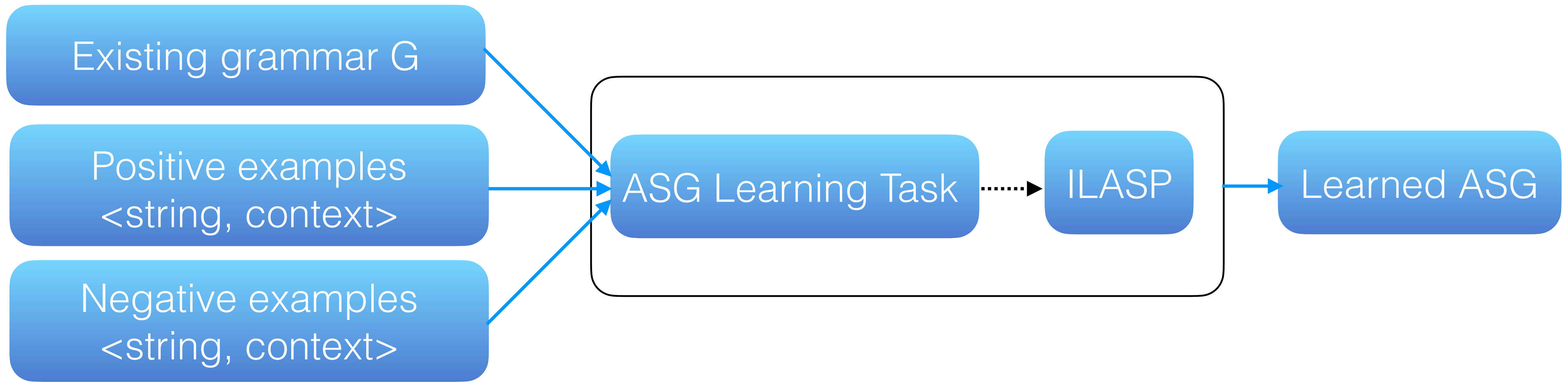
✓ "abc" is accepted by $\mathcal{L}(G)$

✓ "$a^nb^nc^n$" is accepted by $\mathcal{L}(G)$

✗ "ac" is not accepted by $\mathcal{L}(G)$

Learn a class of context-sensitive grammars (ASG):

- context-free part defines the syntax of the language
- context-sensitive parts defines semantics.



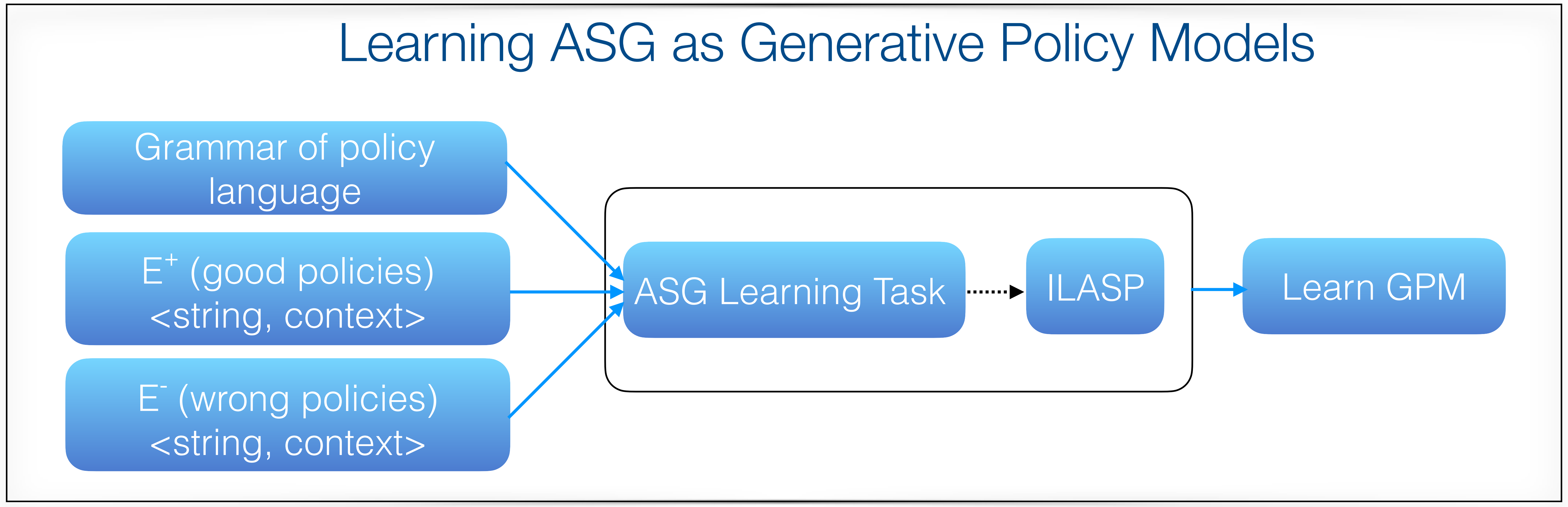
Answer Set Grammars

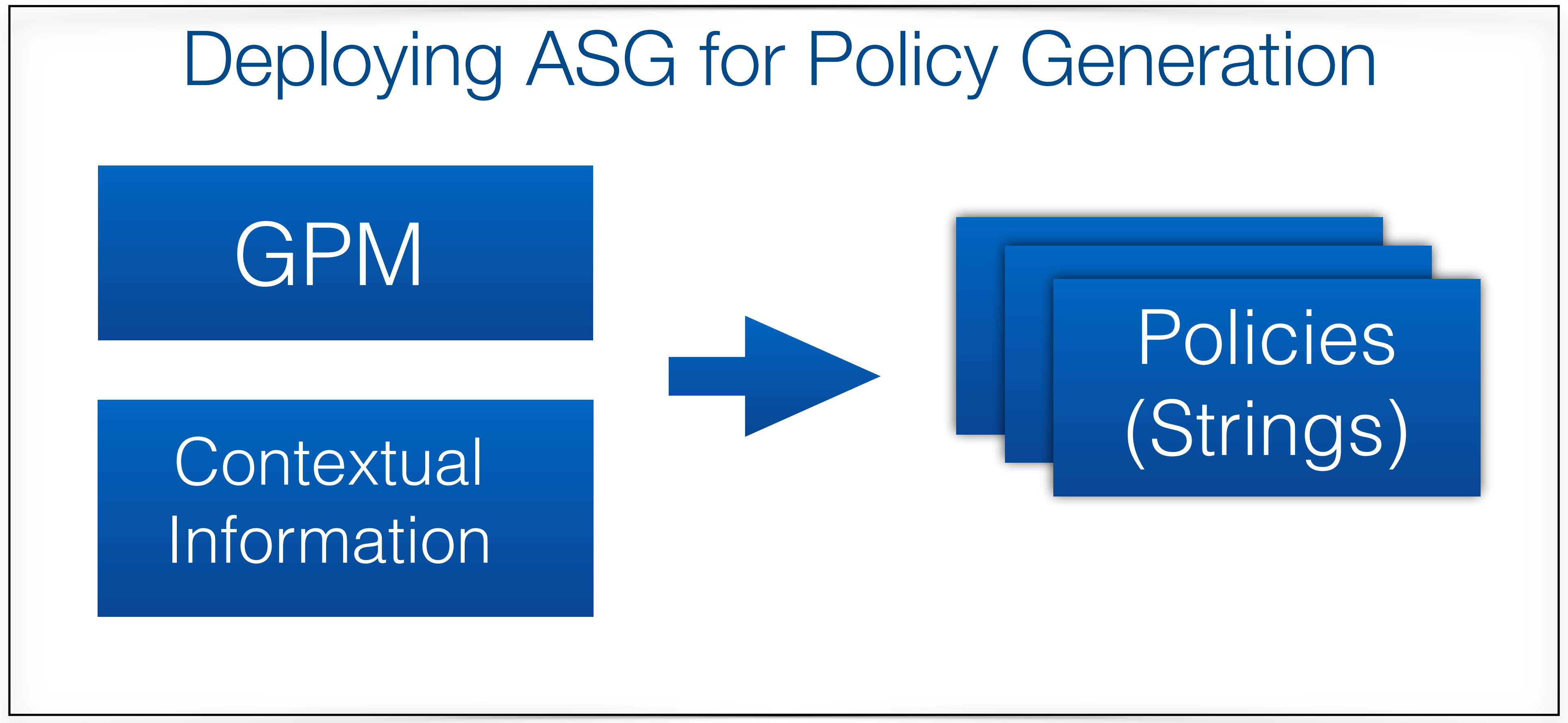
M. Law, A. Russo, E. Bertino, K.Broda, K. Lobo. Representing and Learning Grammars in Answer Set Programming, AAAI19.

Imperial College London

Learning Generative Policy Models

- Intelligent devices/systems need to self-configure to adapt their behaviour in dynamic and complex contexts.
- Generative Policy Model (GPM): a solution for automatic, context-aware generation of policies





Applications

- Autonomous vehicle scenario
- Learning access control policies
- Logistic resupply scenario

Summary of SOTA of Symbolic Learners

System	Normal Rules	Constraints	Non-determinism	Preferences	Context	Noise	Optimal
ASPAL	✓	✗	✗	✗	✗	✗	✓
XHAIL	✓	✗	✗	✗	✗	✓	✗
ILED	✓	✗	✗	✗	✗	✗	✗
OLED	✓	✗	✗	✗	✗	✓	✗
Inspire	✓	✗	✗	✗	✗	✓	✗
ILASP	✓	✓	✓	✓	✓	✓	✓

Imperial College London

Summary of SOTA of Symbolic Learners

System	Normal Rules	Constraints	Non-determinism	Preferences	Context	Noise	Optimal
ASPAL	✓	✗	✗	✗	✗	✗	✓
XHAIL	✓	✗	✗	✗	✗	✓	✗
ILED	✓	✗	✗	✗	✗	✗	✗
OLED	✓	✗	✗	✗	✗	✓	✗
Inspire	✓	✗	✗	✗	✗	✓	✗
ILASP	✓	✓	✓	✓	✓	✓	✓

Two Misconceptions Resolved:

- Complex models expressing recursive concepts, non-monotonic assumptions, constraints, preferences, can be efficiently learned by ILASP.
- ILASP is robust to noise in the data.

Summary of SOTA of Symbolic Learners

System	Normal Rules	Constraints	Non-determinism	Preferences	Context	Noise	Optimal
ASPAL	✓	✗	✗	✗	✗	✗	✓
XHAIL	✓	✗	✗	✗	✗	✓	✗
ILED	✓	✗	✗	✗	✗	✗	✗
OLED	✓	✗	✗	✗	✗	✓	✗
Inspire	✓	✗	✗	✗	✗	✓	✗
ILASP	✓	✓	✓	✓	✓	✓	✓

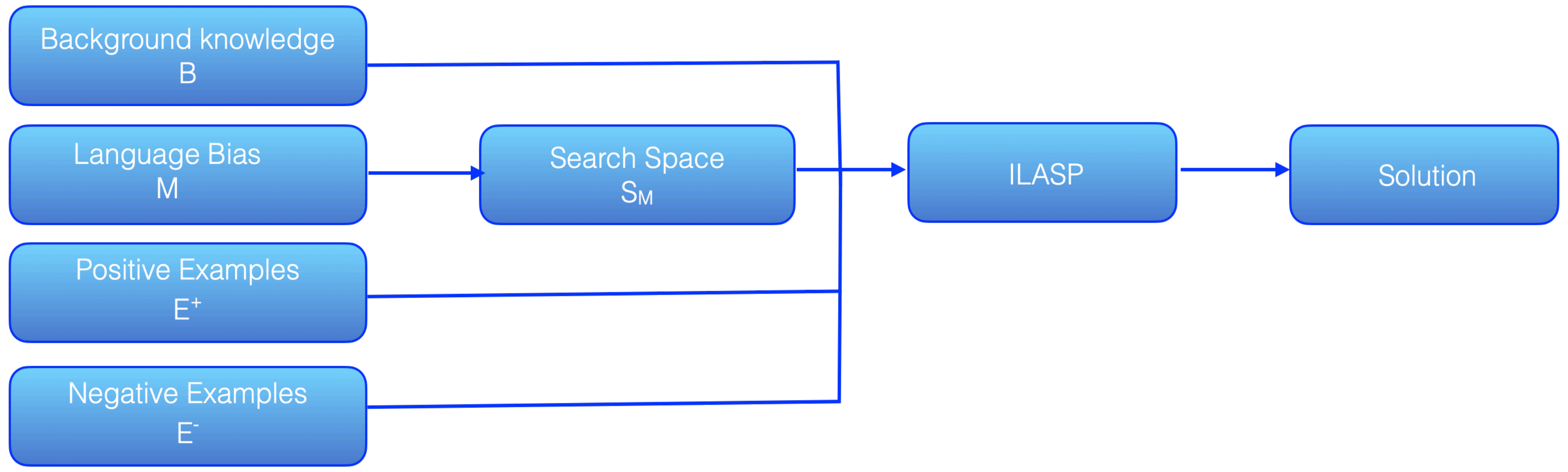
Two Misconceptions Resolved:

- Complex models expressing recursive concepts, non-monotonic assumptions, constraints, preferences, can be efficiently learned by ILASP.
- ILASP is robust to noise in the data.

What about scalability?

FastLAS: Scalable Symbolic Learner

ILASP



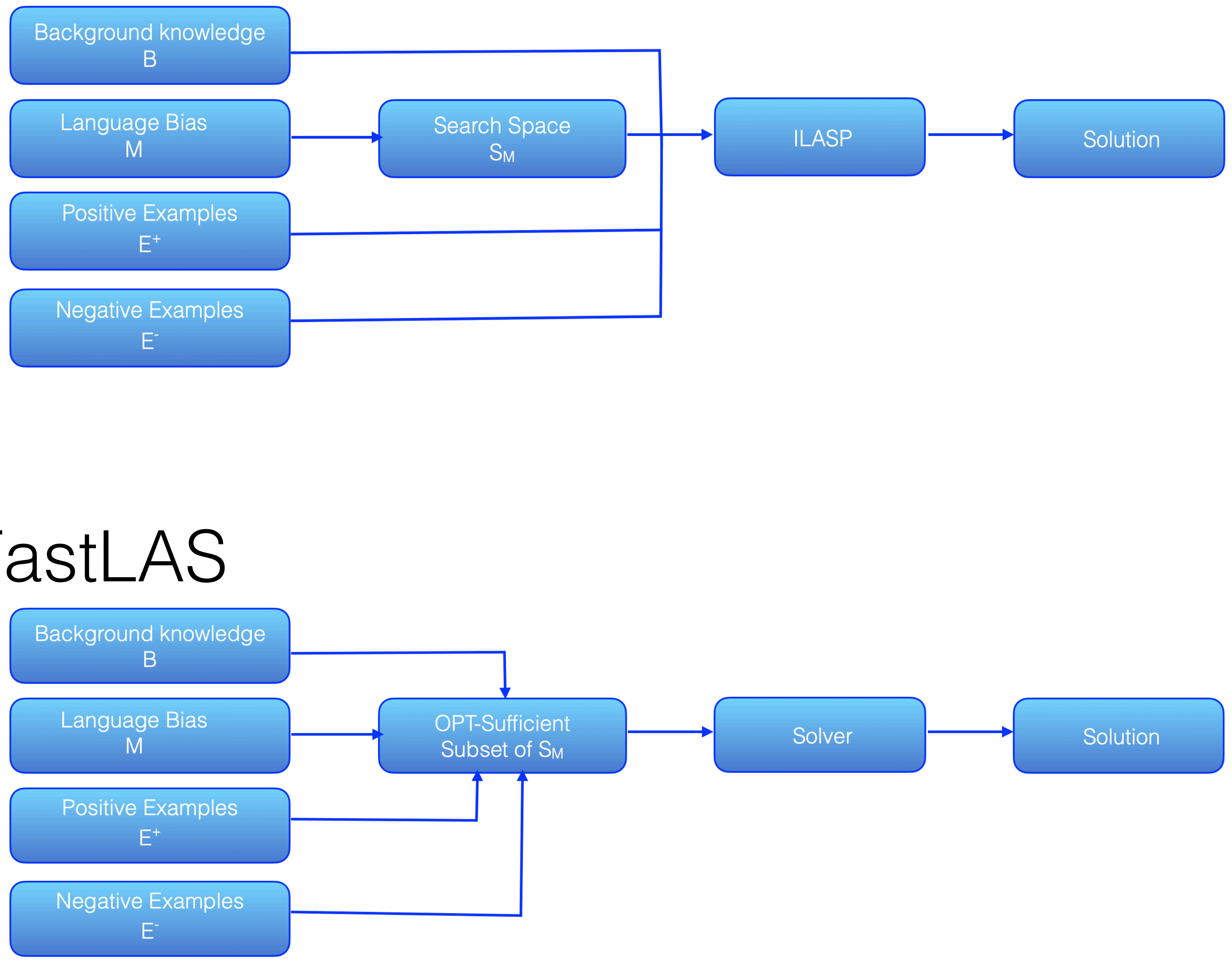
M. Law, A. Russo, E. Bertino, K.Broda, K. Lobo. Representing and Learning Grammars in Answer Set Programming, AAAI2020.

Imperial College London

FastLAS: Scalable Symbolic Learner

ILASP

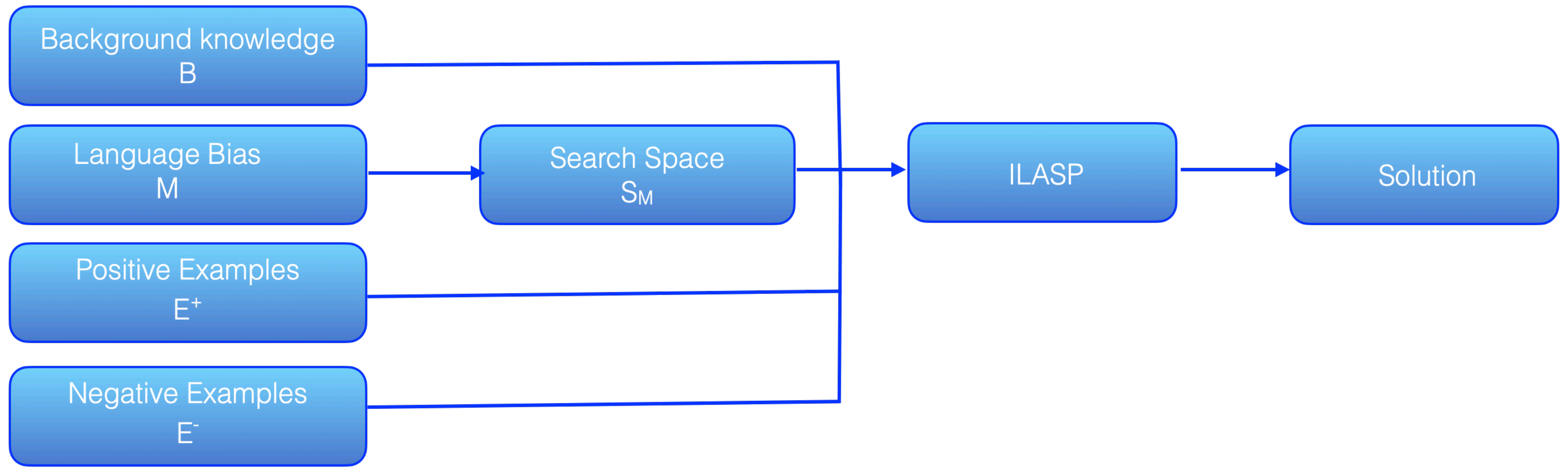
FastLAS



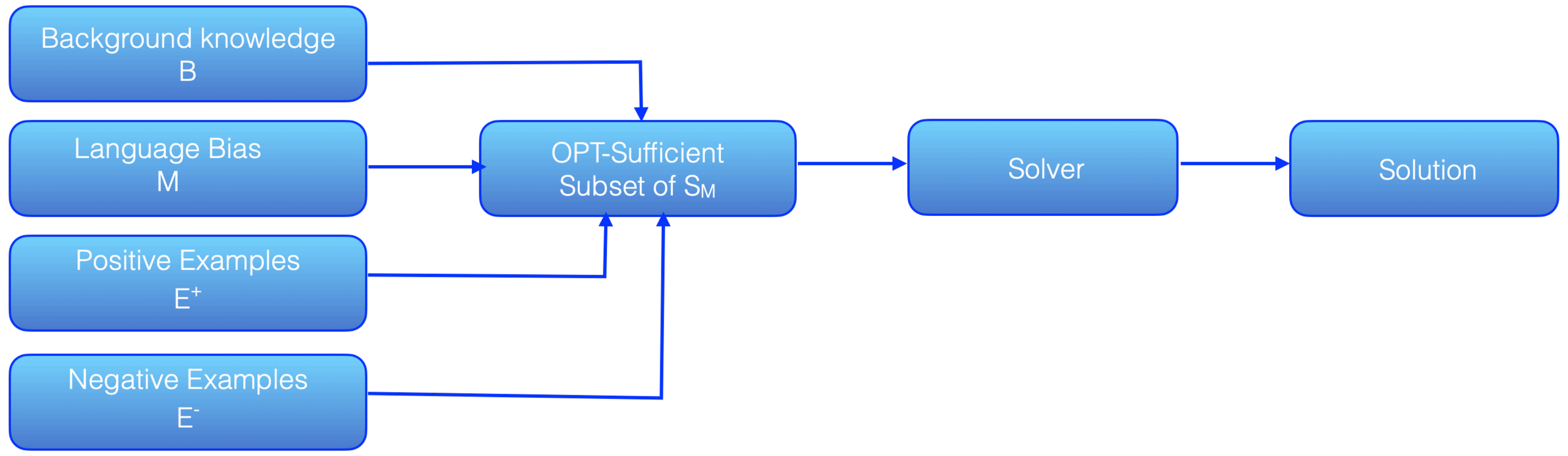
M. Law, A. Russo, E. Bertino, K.Broda, K. Lobo. Representing and Learning Grammars in Answer Set Programming, AAAI2020.

FastLAS: Scalable Symbolic Learner

ILASP



FastLAS



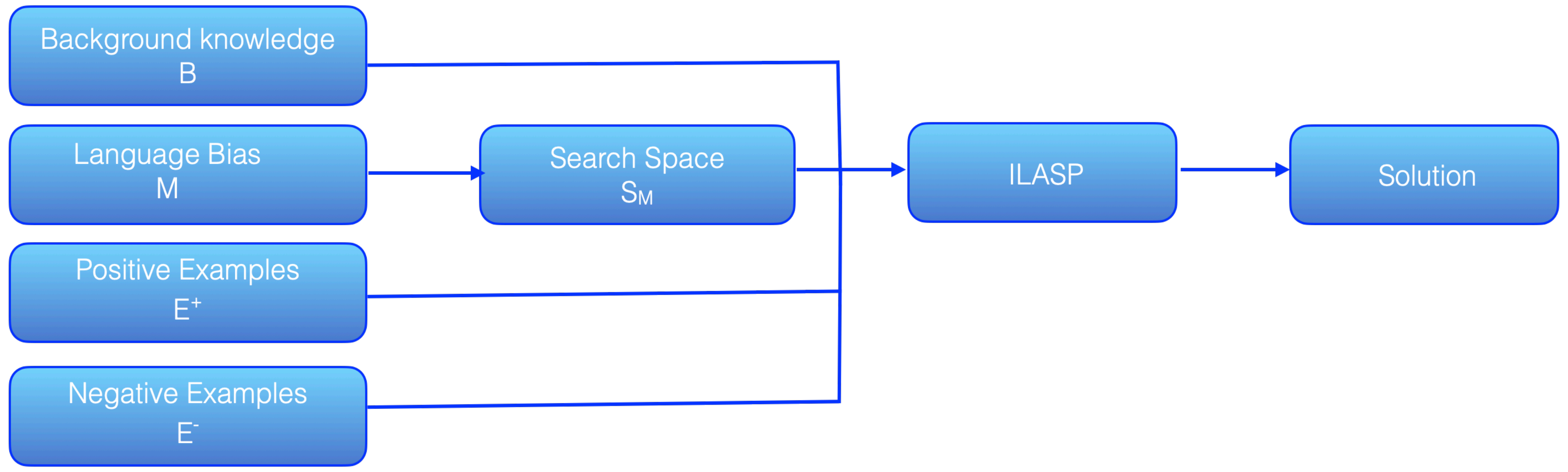
Event detection - CAVIAR dataset

Learn a model that detects people meeting

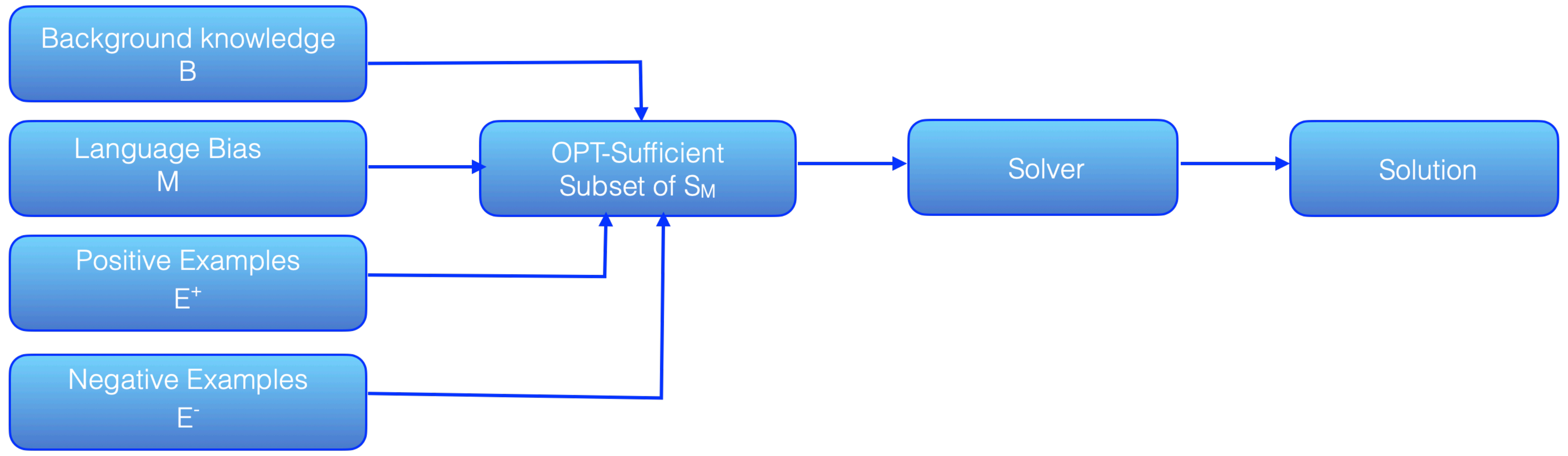
M. Law, A. Russo, E. Bertino, K.Broda, K. Lobo. Representing and Learning Grammars in Answer Set Programming, AAAI2020.

FastLAS: Scalable Symbolic Learner

ILASP



FastLAS



Event detection - CAVIAR dataset

Learn a model that detects people meeting

System	F_1	Running Time
OLED	0.792	107s
ILASP3	0.837	523.3s
FastLAS	0.907	263.8s

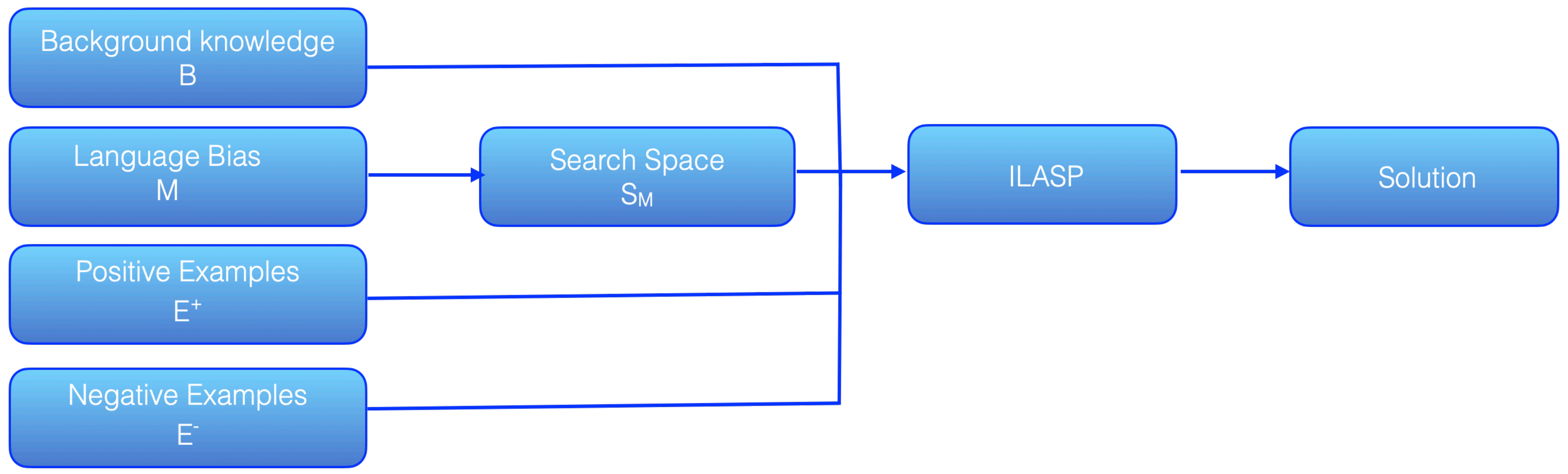
S_M = 3370 rules

$S_M = 2^{44}$ rules

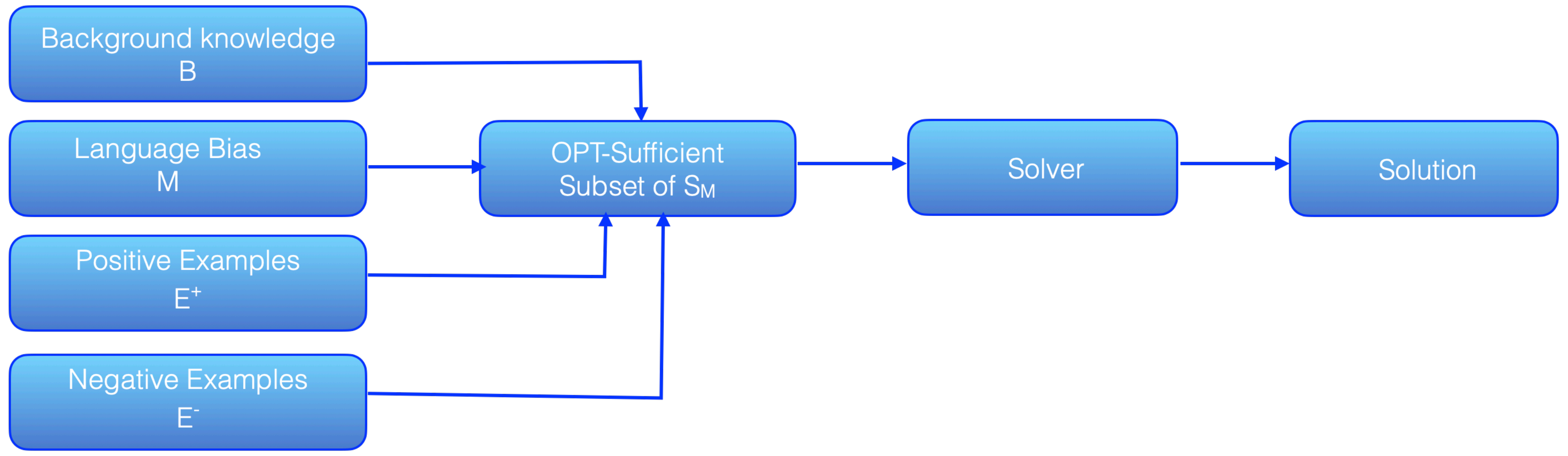
M. Law, A. Russo, E. Bertino, K.Broda, K. Lobo. Representing and Learning Grammars in Answer Set Programming, AAAI2020.

FastLAS: Scalable Symbolic Learner

ILASP



FastLAS



Event detection - CAVIAR dataset

Low-level features (e.g. people's location) already extracted

Learn a model that detects people meeting

System	F_1	Running Time
OLED	0.792	107s
ILASP3	0.837	523.3s
FastLAS	0.907	263.8s

S_M = 3370 rules

$S_M = 2^{44}$ rules

What about if data are unstructured?

M. Law, A. Russo, E. Bertino, K.Broda, K. Lobo. Representing and Learning Grammars in Answer Set Programming, AAAI2020.

Machine Comprehension of Text

Facebook's bAbI dataset

Story:

1. John went to the local restaurant.
2. The waiter brought John a glass of water and took the order.
3. As John was waiting, he took out the book and began to read it.
4. The steak which he ordered finally arrived.
5. After John had finished the meal, he took the jacket but he forgot to take the book.
6. He paid the bill and went back to the hotel.

Questions:

1. Where is John?
2. Where was John before he went to the hotel?
3. Who took the order?
4. Who received a glass of water?
5. Did John have to wait?
6. What did John read?
7. What did John choose?
8. Where is the book?
9. Where is the jacket?

P. Chabierski, A. Russo, M. Law, K. Broda, Machine Comprehension of Text Using Combinatory Categorial Grammar and Answer Set Programs. COMMONSENSE 2017

Imperial College London

Machine Comprehension of Text

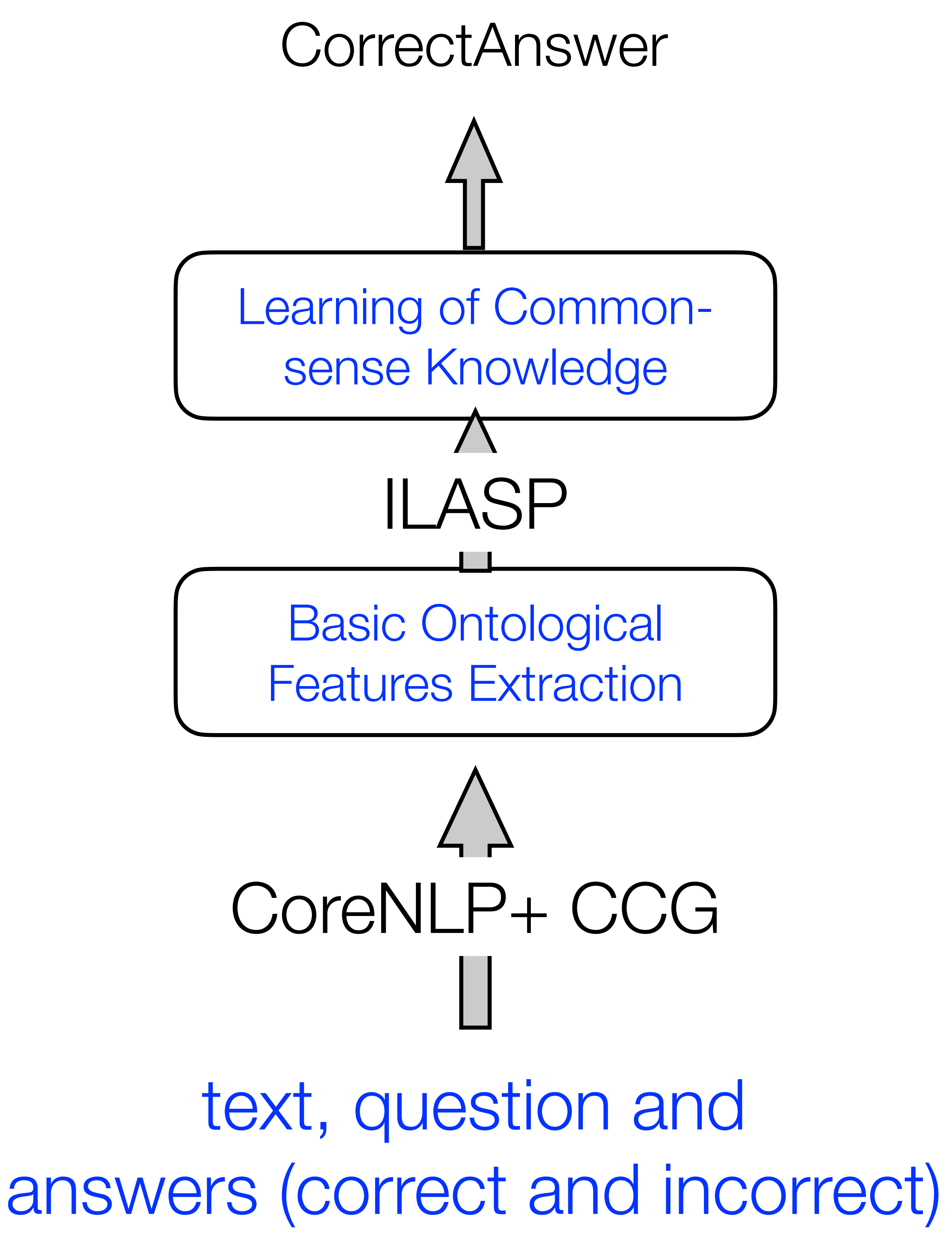
Facebook's bAbI dataset

Story:

1. John went to the local restaurant.
2. The waiter brought John a glass of water and took the order.
3. As John was waiting, he took out the book and began to read it.
4. The steak which he ordered finally arrived.
5. After John had finished the meal, he took the jacket but he forgot to take the book.
6. He paid the bill and went back to the hotel.

Questions:

1. Where is John?
2. Where was John before he went to the hotel?
3. Who took the order?
4. Who received a glass of water?
5. Did John have to wait?
6. What did John read?
7. What did John choose?
8. Where is the book?
9. Where is the jacket?



P. Chabierski, A. Russo, M. Law, K. Broda, Machine Comprehension of Text Using Combinatory Categorial Grammar and Answer Set Programs. COMMONSENSE 2017

Imperial College London

Machine Comprehension of Text

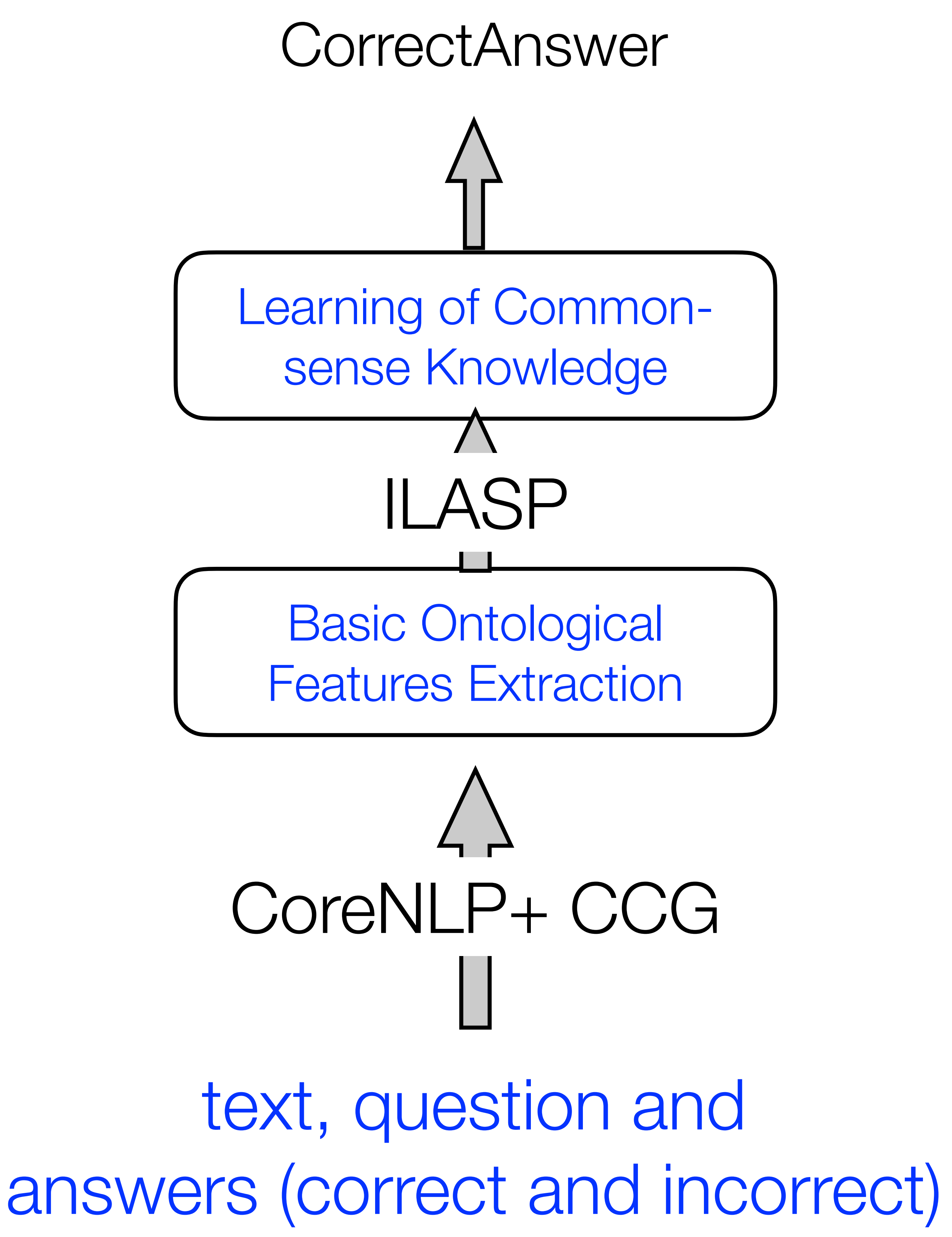
Facebook's bAbI dataset

Story:

1. John went to the local restaurant.
2. The waiter brought John a glass of water and took the order.
3. As John was waiting, he took out the book and began to read it.
4. The steak which he ordered finally arrived.
5. After John had finished the meal, he took the jacket but he forgot to take the book.
6. He paid the bill and went back to the hotel.

Questions:

1. Where is John?
2. Where was John before he went to the hotel?
3. Who took the order?
4. Who received a glass of water?
5. Did John have to wait?
6. What did John read?
7. What did John choose?
8. Where is the book?
9. Where is the jacket?



Automated semantic representation of English text into logic-based knowledge.

P. Chabierski, A. Russo, M. Law, K. Broda, Machine Comprehension of Text Using Combinatory Categorial Grammar and Answer Set Programs. COMMONSENSE 2017

Machine Comprehension of Text

Facebook's bAbI dataset

Story:

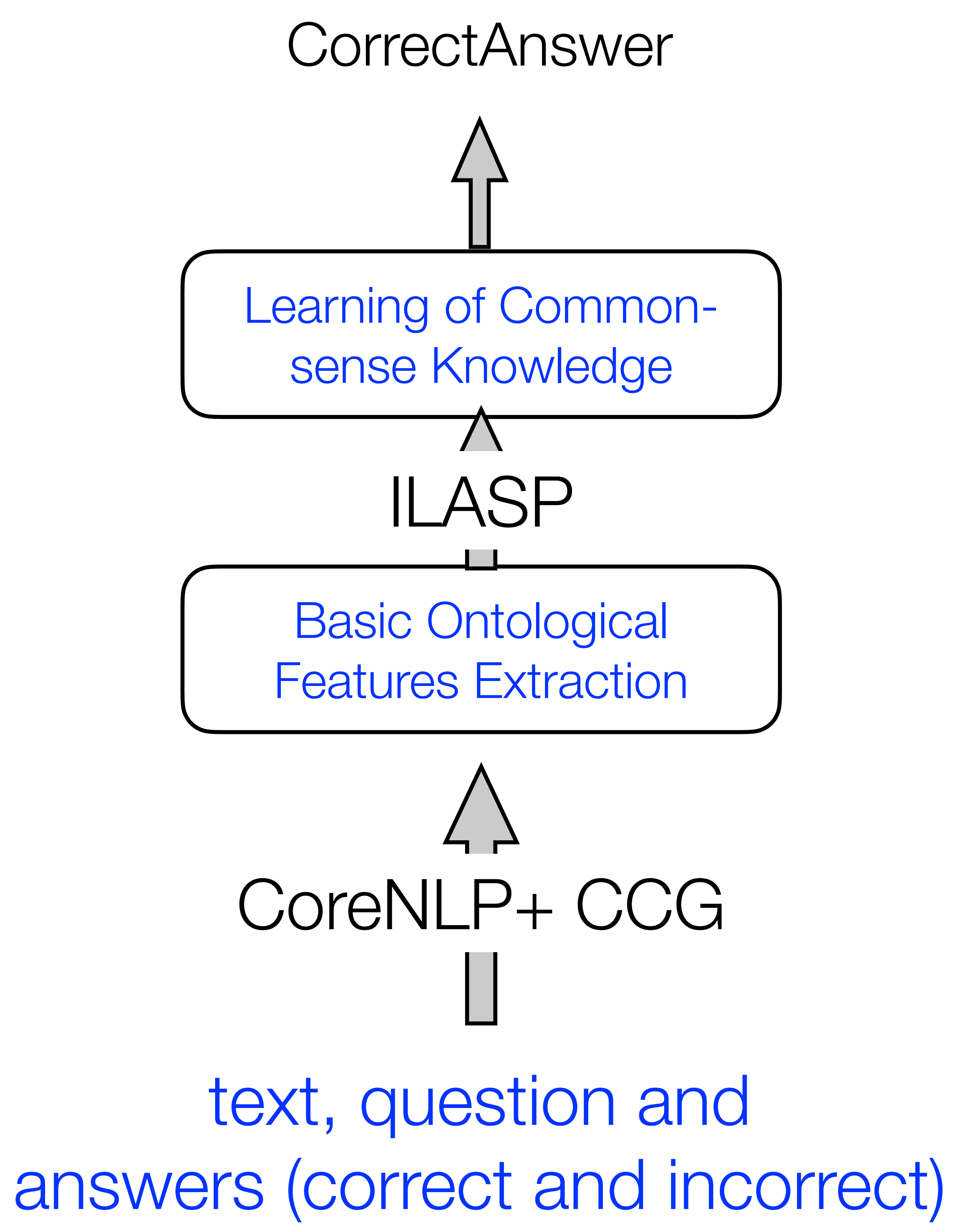
1. John went to the local restaurant.
2. The waiter brought John a glass of water and took the order.
3. As John was waiting, he took out the book and began to read it.
4. The steak which he ordered finally arrived.
5. After John had finished the meal, he took the jacket but he forgot to take the book.
6. He paid the bill and went back to

Questions:

1. Whe
2. Whe the
3. Whe
4. Whe
5. Did
6. Wha
7. Wha
8. Whe
9. Whe

	Task						
Ex.	**1**	**6**	**8**	**9**	**12**	**15**	**18**
5	64.8	90.7	92.4	74.6	100.0	100.0	81.7
10	100.0	90.7	92.4	81.8	100.0	100.0	87.9
15	100.0	92.0	100.0	90.6	100.0	100.0	91.2
20	100.0	95.8	100.0	94.6	100.0	100.0	88.2
25	100.0	98.9	100.0	97.0	100.0	100.0	92.4

	Task						
System	**1**	**6**	**8**	**9**	**12**	**15**	**18**
Sukhbaatar et al. (MemN2N)	99.9	98.0	93.9	98.5	100.0	98.2	90.8
Henaff et al. (EntNet)	99.3	70.0	80.8	68.5	99.2	42.2	91.2
Report (CCG + ILASP)	**100.0**	**98.9**	**100.0**	**97.0**	**100.0**	**100.0**	**92.4**



Automated semantic representation of English text into logic-based knowledge.

P. Chabierski, A. Russo, M. Law, K. Broda, Machine Comprehension of Text Using Combinatory Categorial Grammar and Answer Set Programs. COMMONSENSE 2017

Imperial College London

Next Step and Open Challenges

FastLAS

- ✓ Orders of magnitude faster that state-of-the-art ILASP
- ✓ Can solve machine learning tasks with much larger search spaces
- ✓ Sound and complete - guaranteed to find an optimal solution

Next Step and Open Challenges

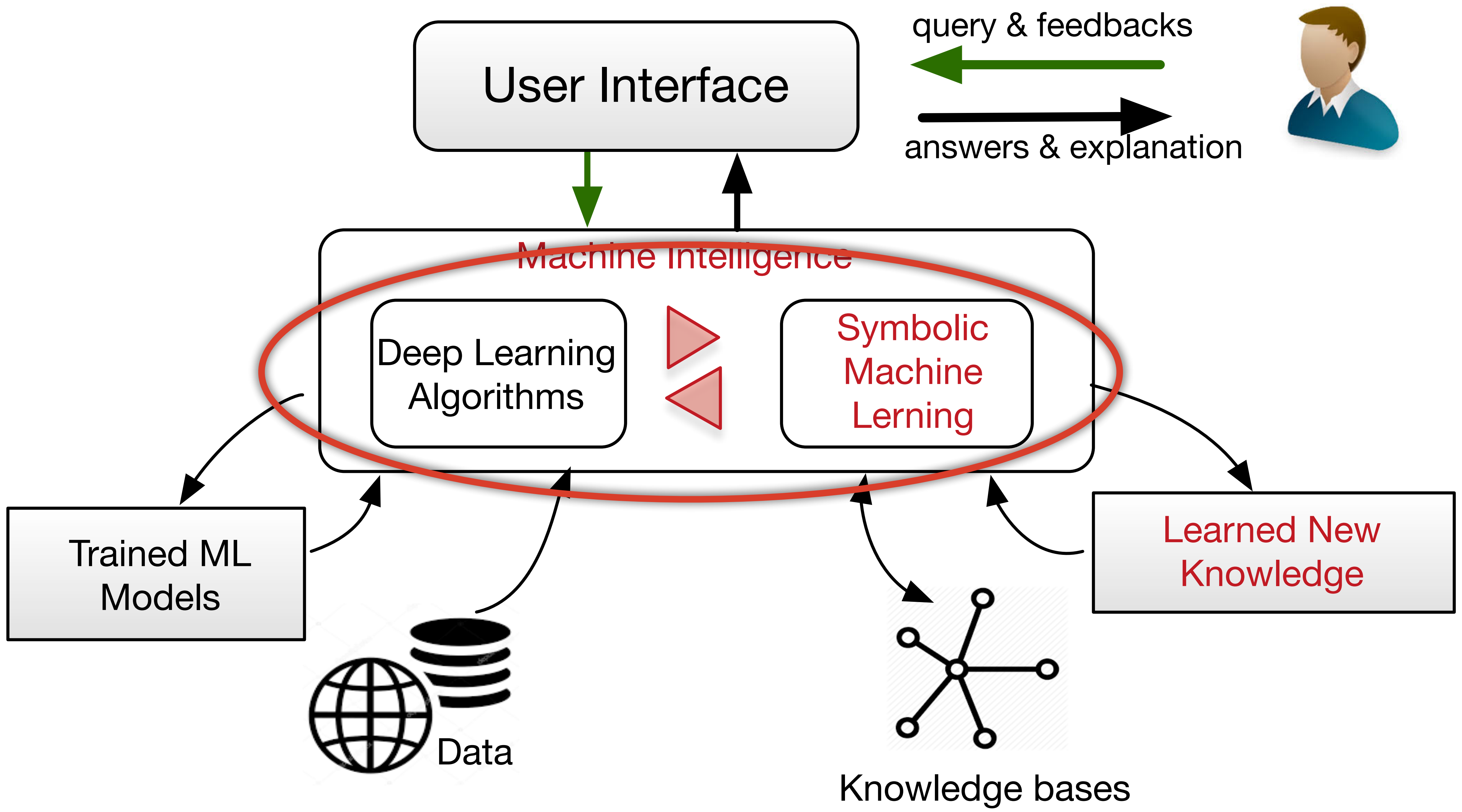
FastLAS

- ✓ Orders of magnitude faster that state-of-the-art ILASP
- ✓ Can solve machine learning tasks with much larger search spaces
- ✓ Sound and complete - guaranteed to find an optimal solution
- ✗ Requires structured data

Next Step and Open Challenges

FastLAS

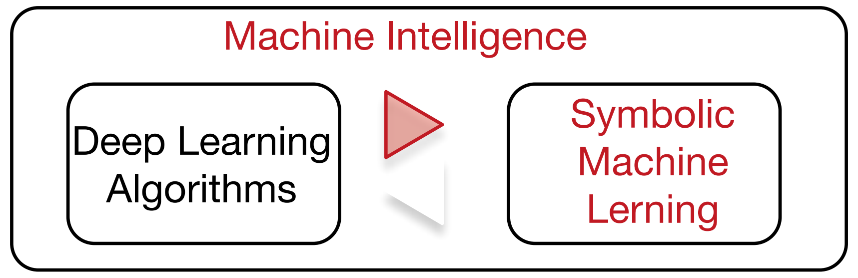
- ✓ Orders of magnitude faster that state-of-the-art ILASP
- ✓ Can solve machine learning tasks with much larger search spaces
- ✓ Sound and complete - guaranteed to find an optimal solution
- ✗ Requires structured data



Imperial College
London

Next Step and Open Challenges

How effective is symbolic rule learning from labelled unstructured data, when contextual information is extracted by deep neural networks?



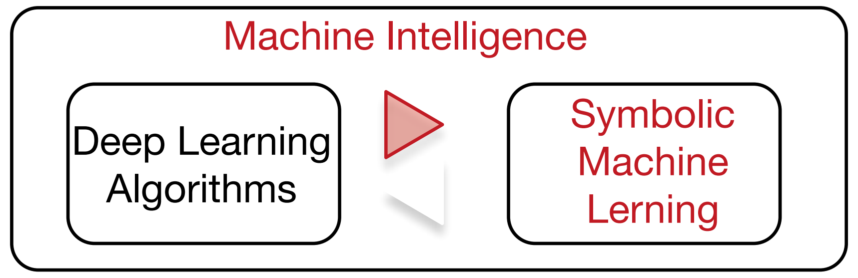
Hybrid Interpretable Learning from Noisy Raw Data

[dstl] DEVCOM ARMY RESEARCH LABORATORY DAIS ITA

D. Cunnington, A. Russo, M. Law, J. Lobo, L. Kaplan. NSL: Hybrid Interpretable Learning From Noisy Raw Data, https://arxiv.org/abs/2012.05023

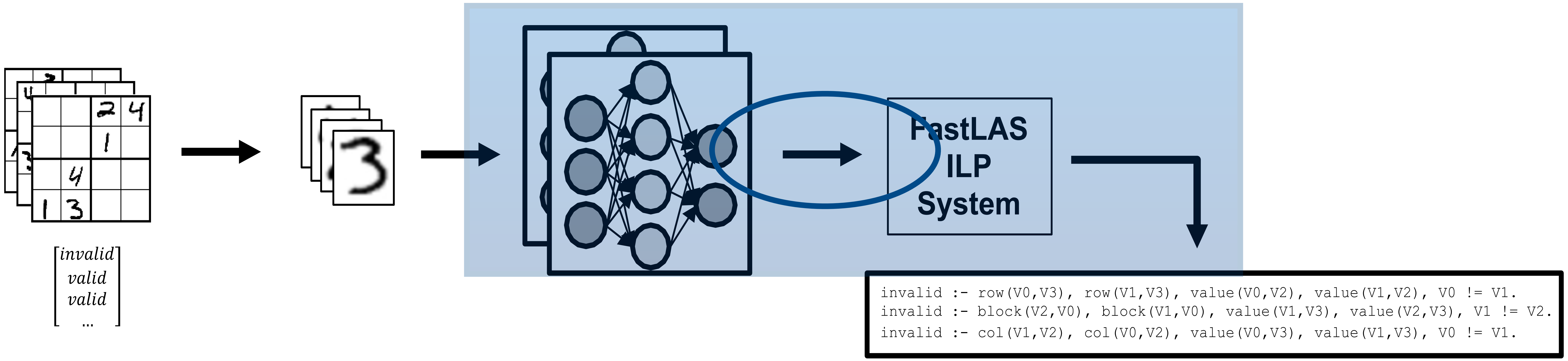
Imperial College
London

Next Step and Open Challenges



How effective is symbolic rule learning from labelled unstructured data, when contextual information is extracted by deep neural networks?

Hybrid Interpretable Learning from Noisy Raw Data



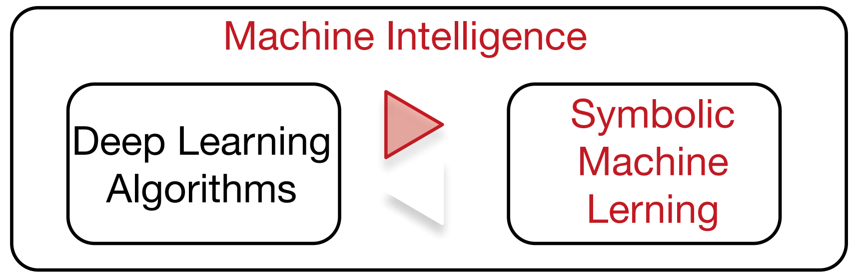
[dstl] DEVCOM ARMY RESEARCH LABORATORY DAIS ITA

D. Cunnington, A. Russo, M. Law, J. Lobo, L. Kaplan. NSL: Hybrid Interpretable Learning From Noisy Raw Data, https://arxiv.org/abs/2012.05023

Imperial College
London

Some Results…

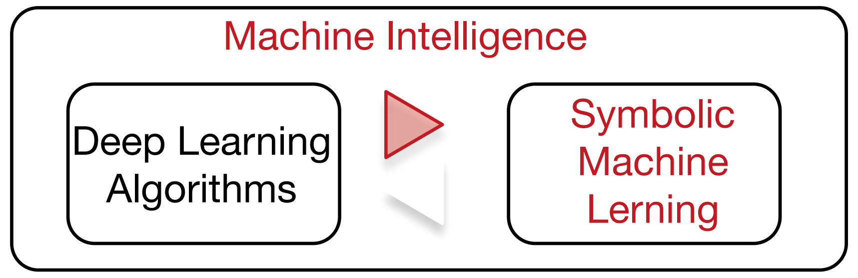
How effective is symbolic rule learning from labelled unstructured data, when contextual information is extracted by deep neural networks?



Hybrid Interpretable Learning from Noisy Raw Data

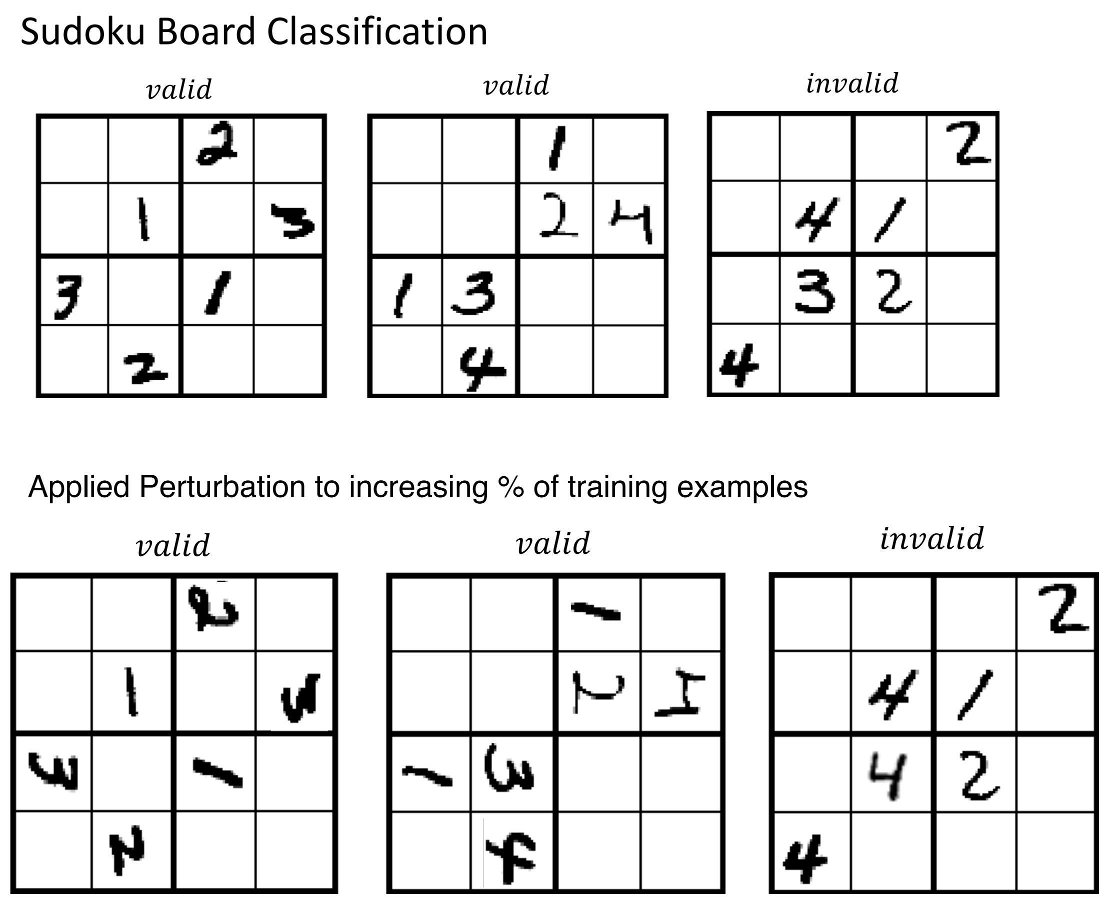
D. Cunnington, A. Russo, M. Law, J. Lobo, L. Kaplan. NSL: Hybrid Interpretable Learning From Noisy Raw Data, https://arxiv.org/abs/2012.05023

Some Results…



How effective is symbolic rule learning from labelled unstructured data, when contextual information is extracted by deep neural networks?

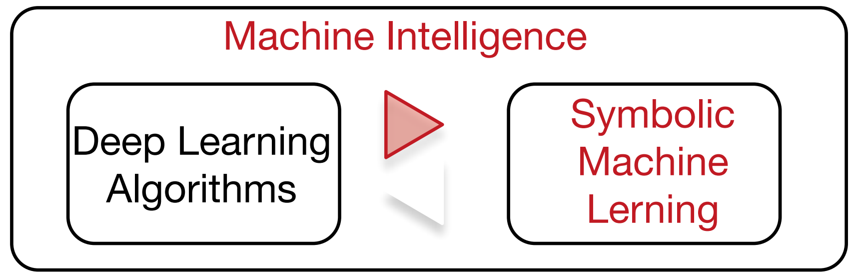
Hybrid Interpretable Learning from Noisy Raw Data



400 examples total
5-fold cross validation:
Train split: 320 examples
Test split: 80 examples

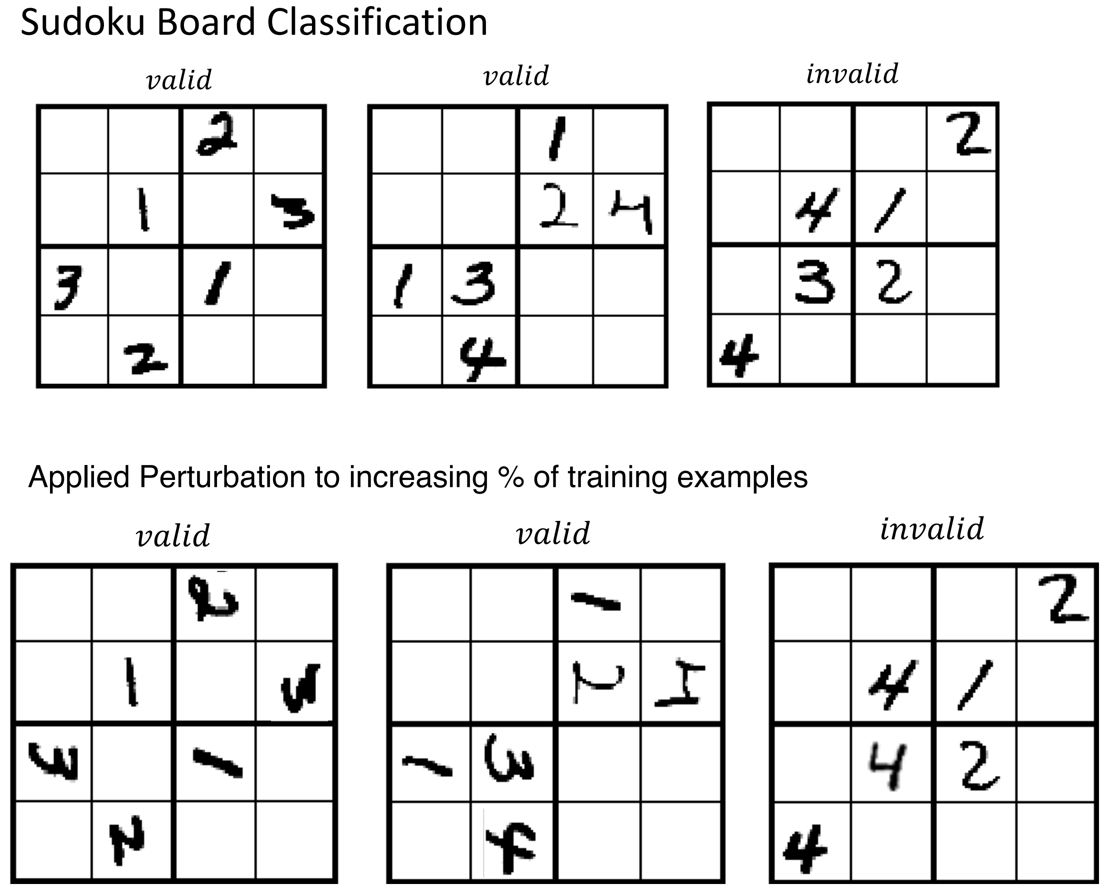
D. Cunnington, A. Russo, M. Law, J. Lobo, L. Kaplan. NSL: Hybrid Interpretable Learning From Noisy Raw Data, https://arxiv.org/abs/2012.05023

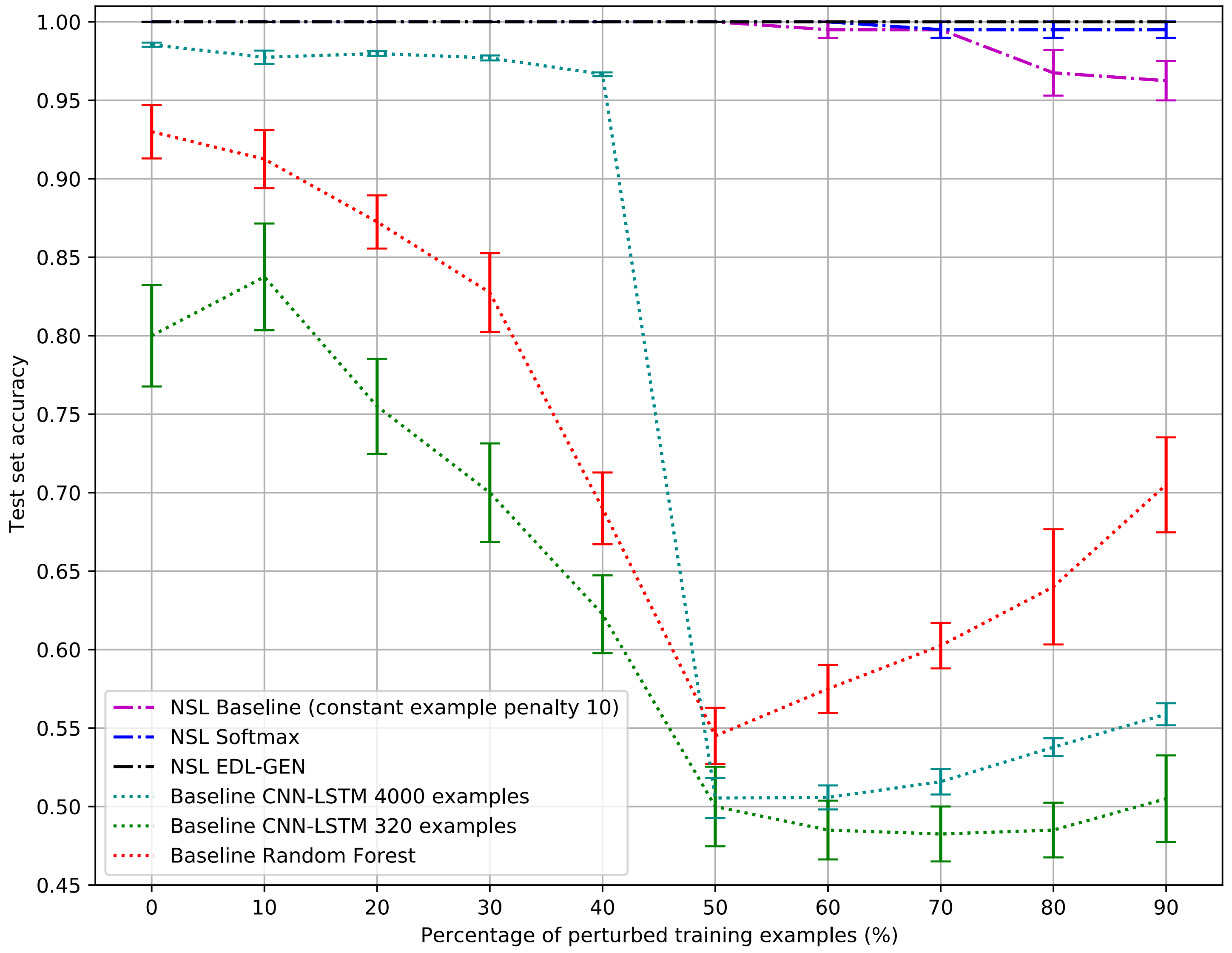
Some Results…



How effective is symbolic rule learning from labelled unstructured data, when contextual information is extracted by deep neural networks?

Hybrid Interpretable Learning from Noisy Raw Data

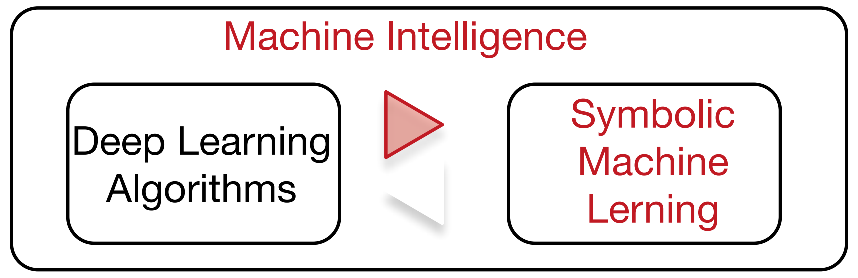




400 examples total
5-fold cross validation:
Train split: 320 examples
Test split: 80 examples

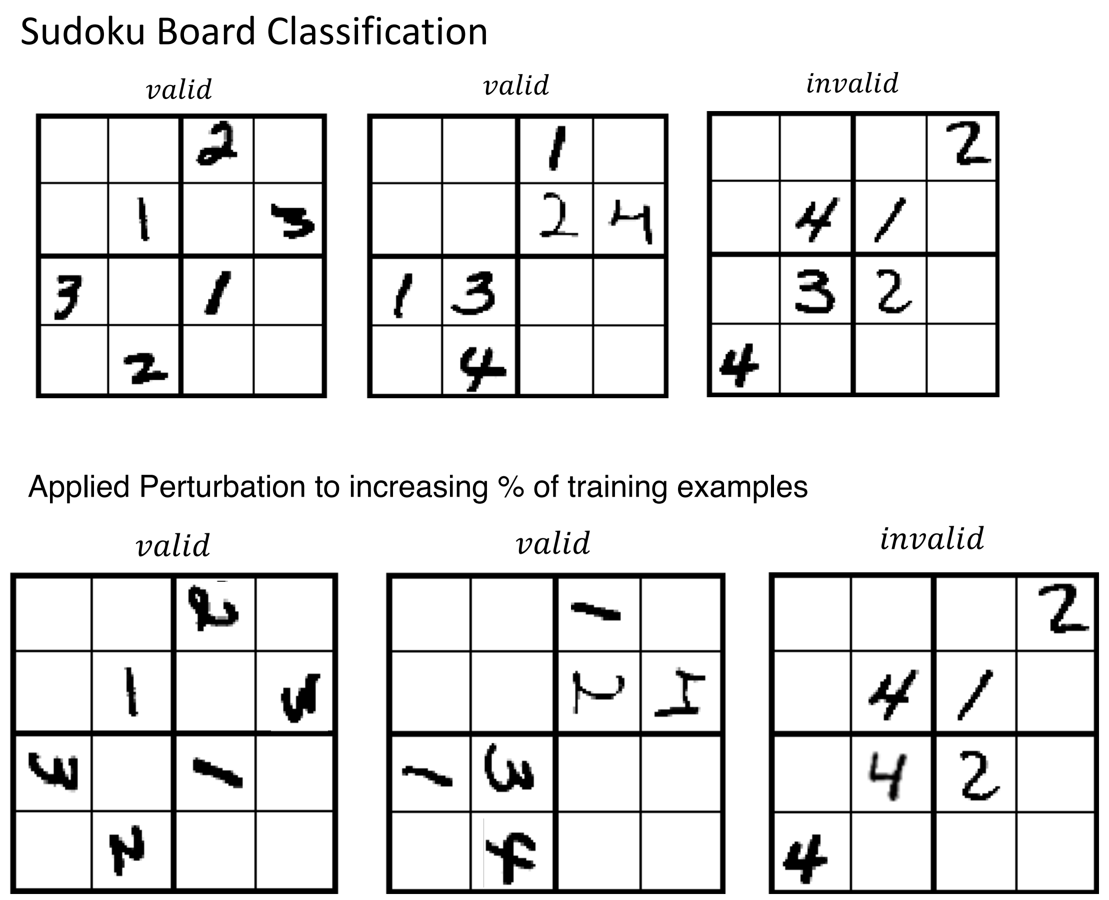
D. Cunnington, A. Russo, M. Law, J. Lobo, L. Kaplan. NSL: Hybrid Interpretable Learning From Noisy Raw Data, https://arxiv.org/abs/2012.05023

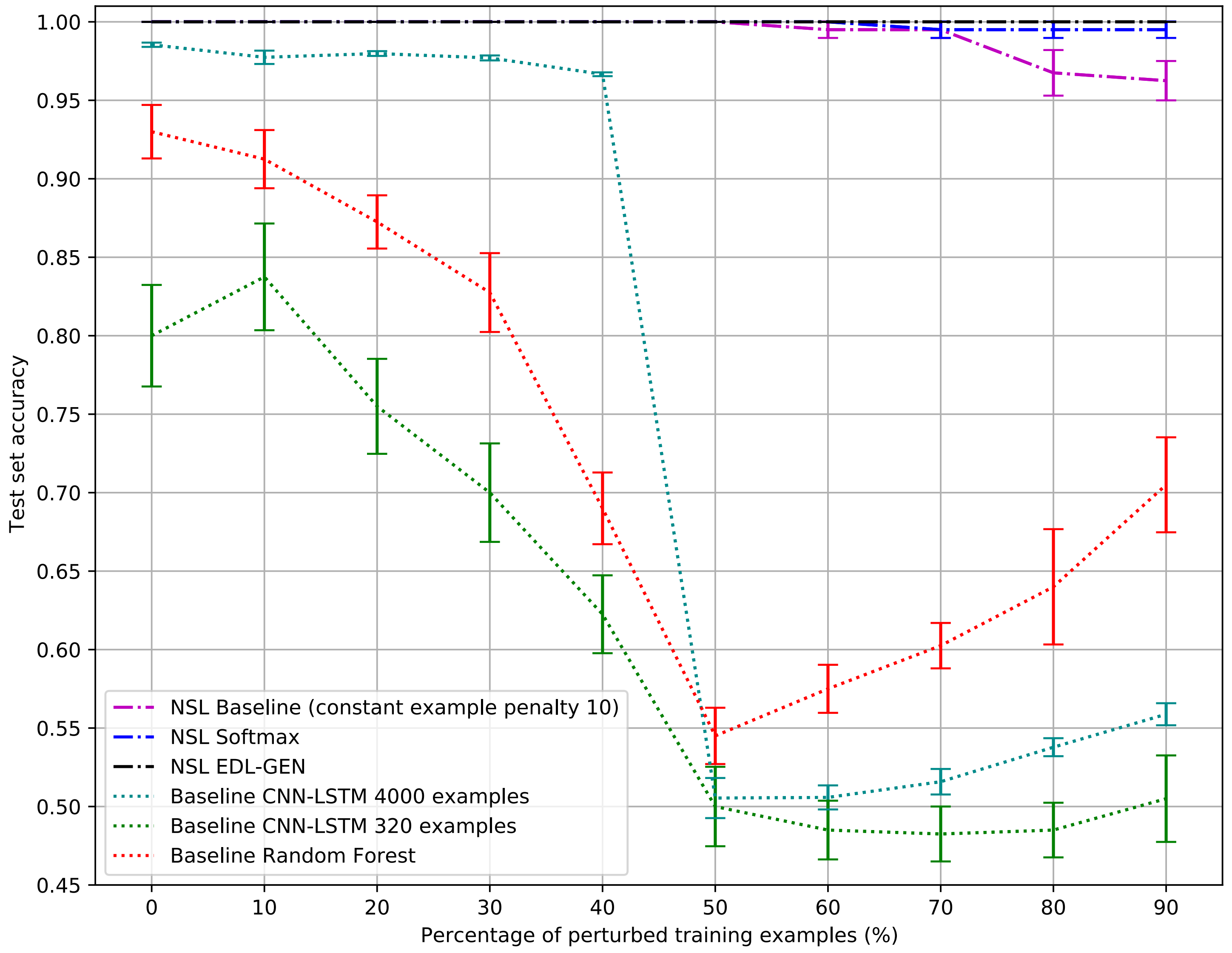
Some Results…

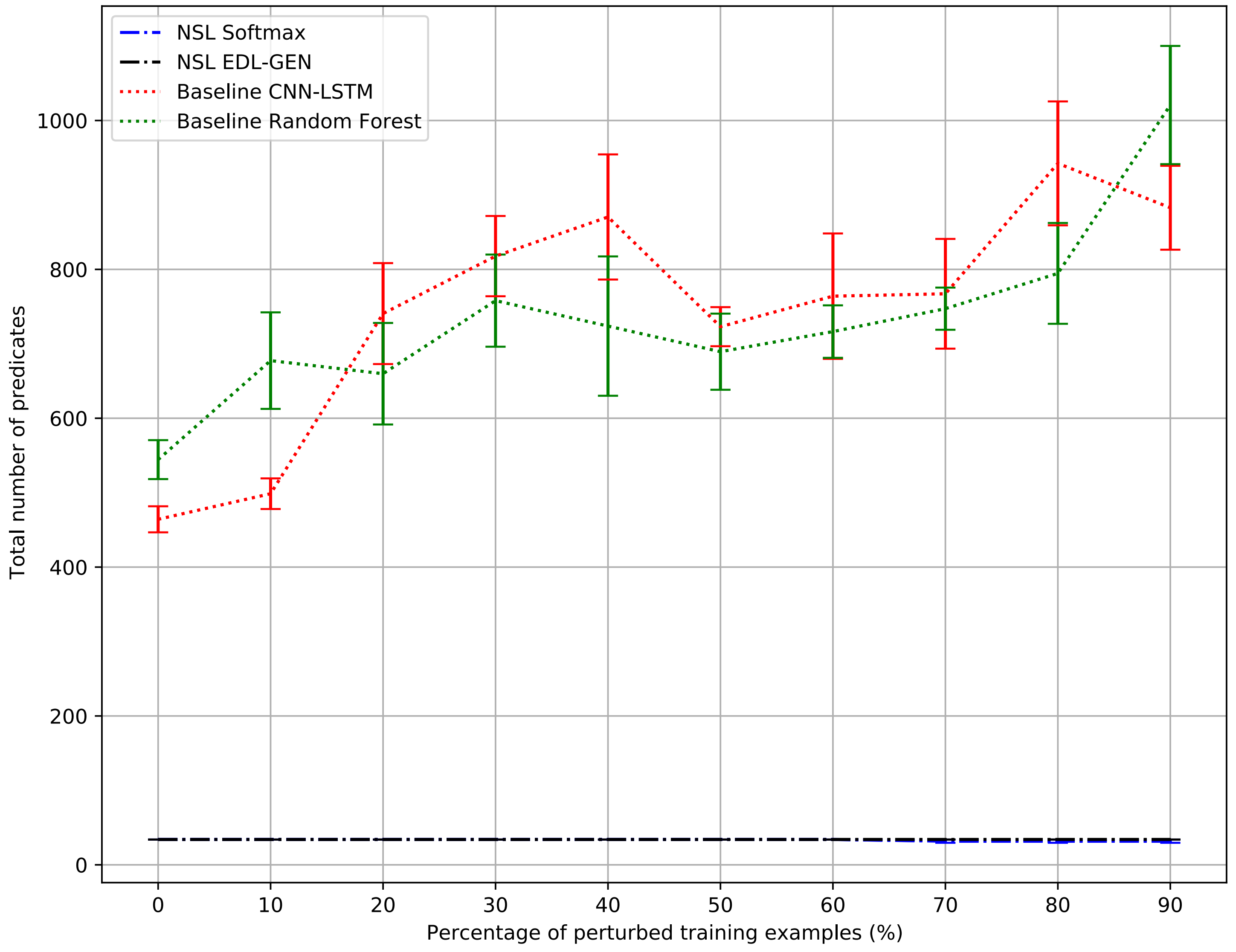


How effective is symbolic rule learning from labelled unstructured data, when contextual information is extracted by deep neural networks?

Hybrid Interpretable Learning from Noisy Raw Data







400 examples total
5-fold cross validation:
Train split: 320 examples
Test split: 80 examples

D. Cunnington, A. Russo, M. Law, J. Lobo, L. Kaplan. NSL: Hybrid Interpretable Learning From Noisy Raw Data, https://arxiv.org/abs/2012.05023

In Summary

Symbolic Machine Learning is capable of

- Learning complex knowledge that expresses recursive concepts, non-monotonic conditions, constraints, preferences.
- Learning generalisations from noisy data without overfitting the data.
- Learning knowledge that is interpretable and that can be used to automatically generate explanations.
- Learning from unstructured data if integrated with sub-symbolic methods.

More needs to be done to:

- Handle uncertainty (if any) during the learning process and quantify the level of uncertainty of predictions.
- Realise an end-to-end neural-symbolic architecture.

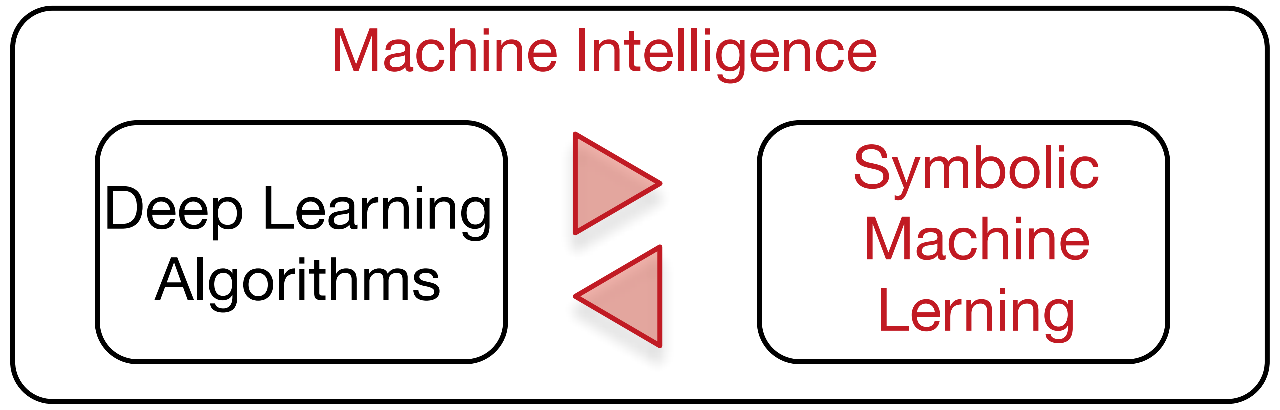
In Summary

Symbolic Machine Learning is capable of

- Learning complex knowledge that expresses recursive concepts, non-monotonic conditions, constraints, preferences.
- Learning generalisations from noisy data without overfitting the data.
- Learning knowledge that is interpretable and that can be used to automatically generate explanations.
- Learning from unstructured data if integrated with sub-symbolic methods.

More needs to be done to:

- Handle uncertainty (if any) during the learning process and quantify the level of uncertainty of predictions.
- Realise an end-to-end neural-symbolic architecture.



Collaborators…

Krysia Broda

Domenico Corapi

Mark Law

Piotr Chabierski

Daniel Cunnington

Jorge Lobo

Elisa Bertino

Collaborators…


Krysia Broda


Domenico Corapi


Mark Law


Piotr Chabierski


Daniel Cunnington


Jorge Lobo


Elisa Bertino

Thank you for listening!

Any questions?