Abduction-Based Explanations for Machine Learning Models

Alexey Ignatiev¹, Nina Narodytska², Joao Marques-Silva¹

January 30, 2019

¹ Faculty of Science, University of Lisbon, Portugal
² VMWare Research, CA, USA
What is eXplainable AI (XAI)?

This is a cat:
- It has fur, whiskers, and claws.
- It has this feature:

Current Explanation

XAI Explanation
Why XAI?

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 27 April 2016
on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

(Text with EEA relevance)
Why XAI?

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 27 April 2016
on the protection of natural persons with regard to the processing of personal data and on the free
movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
(Text with EEA relevance)

European Union regulations on algorithmic decision-making
and a “right to explanation”

Bryce Goodman,1* Seth Flaxman,2
European Union regulations on algorithmic decision-making and a “right to explanation”

Bryce Goodman, Sah Flaxman,

We summarize the potential impact that the European Union’s new General Data Protection Regulation will have on the routine use of machine-learning algorithms. Slated to take effect as law across the European Union in 2018, it will place restrictions on automated individual decision making (that is, algorithms that make decisions based on user-level predictors) that “significantly affect” users. When put into practice, the law may also effectively create a right to explanation, whereby a user can ask for an explanation of an algorithmic decision that significantly affects them. We argue that while this law may pose large challenges for industry, it highlights opportunities for computer scientists to take the lead in designing algorithms and evaluation frameworks that avoid discrimination and enable explanation.
REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 27 April 2016

on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

(Text with EEA relevance)

European Union regulations on algorithmic decision-making and a “right to explanation”

Bryce Goodman,1* Seth Flaxman,2

We summarize the potential impact that the European Union’s new General Data Protection Regulation will have on the routine use of machine-learning algorithms. Slated to take effect as law across the European Union in 2018, it will place restrictions on automated individual decision making (that is, algorithms that make decisions based on user-level predictors) that “significantly affect” users. When put into practice, the law may also effectively create a right to explanation, whereby a user can ask for an explanation of an algorithmic decision that significantly affects them. We argue that while this law may pose large challenges for industry, it highlights opportunities for computer scientists to take the lead in designing algorithms and evaluation frameworks that avoid discrimination and enable explanation.

SUMMIT ON MACHINE LEARNING MEETS FORMAL METHODS
We summarize the potential impact that the European Union’s new General Data Protection Regulation will have on the routine use of machine-learning algorithms. Slated to take effect as law across the European Union in 2018, it will place restrictions on automated individual decision making (that is, algorithms that make decisions based on user-level predictors) that “significantly affect” users. When put into practice, the law may also effectively create a right to explanation, whereby a user can ask for an explanation of an algorithmic decision that significantly affects them. We argue that while this law may pose large challenges for industry, it highlights opportunities for computer scientists to take the lead in designing algorithms and evaluation frameworks that avoid discrimination and enable explanation.
XAI controversy

**MIT Technology Review**

The Dark Secret at the Heart of AI
Will Knight
April 11, 2017

**The Wall Street Journal**

Inside DARPA’s Push to Make Artificial Intelligence Explain Itself
Sara Castellanos and Steven Norton
August 10, 2017

**The Register**

You better explain yourself, mister: DARPA’s mission to make an accountable AI
Dan Robinson
September 29, 2017

**The New York Times Magazine**

Can A.I. Be Taught to Explain Itself?
Cliff Kuang
November 21, 2017

**ExecutiveBiz**

Charles River Analytics-Led Team Gets DARPA Contract to Support Artificial Intelligence Program
Ramona Adams
June 13, 2017

**Entrepreneur**

Elon Musk and Mark Zuckerberg Are Arguing About AI -- But They’re Both Missing the Point
Artur Kuliian
July 28, 2017

**Military Embedded Systems**

Team investigates artificial intelligence, machine learning in DARPA project
Lisa Daigle
June 14, 2017

**Fast Company**

Why The Military And Corporate America Want To Make AI Explain Itself
Steven Melendez
June 22, 2017

**Computerworld**

Oracle quietly researching ‘Explainable AI’
George Nota
May 5, 2017

**Scientific American**

Demystifying the Black Box That Is AI
Ariel Bleicher
August 9, 2017

**JANES**

DARPA’s XAI seeks explanations from autonomous systems
Geoff Fein
November 16, 2017

**NOVA NEXT**

Ghosts in the Machine
Christina Couch
October 25, 2017

**Science**

How AI detectives are cracking open the black box of deep learning
Paul Voosen
July 6, 2017
heuristic approaches exist
(e.g. LIME or Anchor)
heuristic approaches exist
(e.g. LIME or Anchor)

• *local* explanations
heuristic approaches exist
(e.g. LIME or Anchor)

• local explanations
• no guarantees
heuristic approaches exist
(e.g. LIME or Anchor)

- local explanations
- no guarantees

(un-)reliable?
From ML model to logic

Machine Learning System
From ML model to logic

Machine Learning System

formula $\mathcal{F}$
From ML model to logic

Machine Learning System

cube $C$

formula $\mathcal{F}$

Cat
From ML model to logic

Machine Learning System

cube $C$

formula $\mathcal{F}$

literal $E$
From ML model to logic

\[ C \wedge \mathcal{F} \models \mathcal{E} \]
given a classifier $\mathcal{F}$, a cube $C$ and a prediction $\mathcal{E}$,
given a \textit{classifier} $\mathcal{F}$, a \textit{cube} $\mathcal{C}$ and a \textit{prediction} $\mathcal{E}$, compute a (\textit{cardinality-} or \textit{subset-}) minimal $\mathcal{C}_m \subseteq \mathcal{C}$ s.t.
Abductive explanations of ML models

given a classifier $\mathcal{F}$, a cube $\mathcal{C}$ and a prediction $\mathcal{E}$, compute a (cardinality- or subset-) minimal $\mathcal{C}_m \subseteq \mathcal{C}$ s.t.

$$\mathcal{C}_m \land \mathcal{F} \not\models \bot$$

and

$$\mathcal{C}_m \land \mathcal{F} \models \mathcal{E}$$
Abductive explanations of ML models

given a \textit{classifier} $\mathcal{F}$, a \textit{cube} $\mathcal{C}$ and a \textit{prediction} $\mathcal{E}$,
compute a (\textit{cardinality-} or \textit{subset-}) minimal $\mathcal{C}_m \subseteq \mathcal{C}$ s.t.

\[ \mathcal{C}_m \land \mathcal{F} \not\models \top \]

and

\[ \mathcal{C}_m \land \mathcal{F} \models \mathcal{E} \]

\textit{iterative explanation procedure}
Computing primes

1. $C_m \land \sum F \neq \bot$
Computing primes

1. $C_m \land F \not\models \bot$ — tautology
Computing primes

1. $C_m \land F \not\models \bot$ — tautology
2. $C_m \land F \models E$
1. $C_m \land \mathcal{F} \not\models \bot$ — tautology

2. $C_m \land \mathcal{F} \models \mathcal{E}$ $\iff$ $C_m \models (\mathcal{F} \rightarrow \mathcal{E})$
1. $C_m \land \mathcal{F} \not\models \bot$ — tautology

2. $C_m \land \mathcal{F} \models \mathcal{E} \iff C_m \models (\mathcal{F} \rightarrow \mathcal{E})$

$C_m$ is a prime implicant of $\mathcal{F} \rightarrow \mathcal{E}$
Computing one subset-minimal explanation

Input: $\mathcal{F}$ under $\mathcal{M}$, initial cube $\mathcal{C}$, prediction $\mathcal{E}$
Output: Subset-minimal explanation $\mathcal{C}_m$

1. begin
2. for $l \in \mathcal{C}$:
3.  if Entails($\mathcal{C} \setminus \{l\}, \mathcal{F} \rightarrow \mathcal{E}, \mathcal{M}$):
4.   $\mathcal{C} \leftarrow \mathcal{C} \setminus \{l\}$
5.  return $\mathcal{C}$
6. end
cardinality-minimal explanations can be computed

(following implicit-hitting set based approach\(^1\))

---

\(^1\)Ignatiev, A.; Morgado, A.; and Marques-Silva, J. 2016. *Propositional abduction with implicit hitting sets*. In ECAI, 1327–1335
Computing one cardinality-minimal explanation

cardinality-minimal explanations can be computed (following *implicit-hitting set* based approach\(^1\))

see the paper

---

\(^1\)Ignatiev, A.; Morgado, A.; and Marques-Silva, J. 2016. *Propositional abduction with implicit hitting sets*. In ECAI, 1327–1335
Experimental setup

- implementation in Python
  - supports SMT solvers through PySMT
    - Yices2 used
  - supports CPLEX 12.8.0

---

Experimental setup

- implementation in Python
  - supports SMT solvers through PySMT
    - Yices2 used
  - supports CPLEX 12.8.0
- \textit{ReLU-based} neural networks\textsuperscript{2}
  - one \textit{hidden} layer with $i \in \{10, 15, 20\}$ neurons

Experimental setup

- implementation in Python
  - supports SMT solvers through PySMT
    - Yices2 used
  - supports CPLEX 12.8.0
- ReLU-based neural networks\(^2\)
  - one hidden layer with \(i \in \{10, 15, 20\}\) neurons
- benchmarks selected from:
  - UCI Machine Learning Repository
  - Penn Machine Learning Benchmarks
  - MNIST Digits Database

Experimental setup

- implementation in Python
  - supports SMT solvers through PySMT
    - Yices2 used
  - supports CPLEX 12.8.0
- ReLU-based neural networks\(^2\)
  - one hidden layer with \(i \in \{10, 15, 20\}\) neurons
- benchmarks selected from:
  - UCI Machine Learning Repository
  - Penn Machine Learning Benchmarks
  - MNIST Digits Database
- Machine configuration:
  - Intel Core i7 2.8GHz, 8GB byte
  - time limit — 1800s
  - memory limit — 4GB byte

### Some of the experimental results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Minimal explanation</th>
<th>Minimum explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>size</td>
<td>SMT (s)</td>
</tr>
<tr>
<td><strong>australian</strong></td>
<td>(14)</td>
<td>&lt;br&gt; m 1</td>
</tr>
<tr>
<td><strong>backache</strong></td>
<td>(32)</td>
<td>&lt;br&gt; m 13</td>
</tr>
<tr>
<td><strong>breast-cancer</strong></td>
<td>(9)</td>
<td>&lt;br&gt; m 3</td>
</tr>
<tr>
<td><strong>cleve</strong></td>
<td>(13)</td>
<td>&lt;br&gt; m 4</td>
</tr>
<tr>
<td><strong>hepatitis</strong></td>
<td>(19)</td>
<td>&lt;br&gt; m 6</td>
</tr>
<tr>
<td><strong>voting</strong></td>
<td>(16)</td>
<td>&lt;br&gt; m 3</td>
</tr>
<tr>
<td><strong>spect</strong></td>
<td>(22)</td>
<td>&lt;br&gt; m 3</td>
</tr>
</tbody>
</table>
Some of the experimental results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Minimal explanation</th>
<th>Minimum explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>size</td>
<td>SMT (s)</td>
</tr>
<tr>
<td>australian</td>
<td>m</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>8.79</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>14</td>
</tr>
<tr>
<td>backache</td>
<td>m</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>19.28</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>26</td>
</tr>
<tr>
<td>breast-cancer</td>
<td>m</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>5.15</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>9</td>
</tr>
<tr>
<td>cleve</td>
<td>m</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>8.62</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>13</td>
</tr>
<tr>
<td>hepatitis</td>
<td>m</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>11.42</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>19</td>
</tr>
<tr>
<td>voting</td>
<td>m</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>4.56</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>11</td>
</tr>
<tr>
<td>spect</td>
<td>m</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>7.31</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>20</td>
</tr>
</tbody>
</table>
### Some of the experimental results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Dataset size</th>
<th>Minimal explanation</th>
<th>Minimum explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>size</td>
<td>SMT (s)</td>
</tr>
<tr>
<td>australian</td>
<td>14</td>
<td>m a M</td>
<td>1 8.79 14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m a M</td>
<td>13 9.28 26</td>
</tr>
<tr>
<td>backache</td>
<td>32</td>
<td>m a M</td>
<td>3 5.15 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m a M</td>
<td>4 8.62 13</td>
</tr>
<tr>
<td>breast-cancer</td>
<td>9</td>
<td>m a M</td>
<td>6 11.42 19</td>
</tr>
<tr>
<td>cleve</td>
<td>13</td>
<td>m a M</td>
<td>3 4.56 11</td>
</tr>
<tr>
<td>hepatitis</td>
<td>19</td>
<td>m a M</td>
<td>3 7.31 20</td>
</tr>
<tr>
<td>voting</td>
<td>16</td>
<td>m a M</td>
<td>— — —</td>
</tr>
</tbody>
</table>
### Some of the experimental results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Minimal explanation</th>
<th></th>
<th>Minimum explanation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>size</td>
<td>SMT (s)</td>
<td>MILP (s)</td>
<td>size</td>
</tr>
<tr>
<td>australian</td>
<td>m</td>
<td>1</td>
<td>0.03</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>8.79</td>
<td>1.38</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>14</td>
<td>17.00</td>
<td>a</td>
</tr>
<tr>
<td>backache</td>
<td>m</td>
<td>13</td>
<td>0.13</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>19.28</td>
<td>5.08</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>26</td>
<td>22.21</td>
<td>a</td>
</tr>
<tr>
<td>breast-cancer</td>
<td>m</td>
<td>3</td>
<td>0.02</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>5.15</td>
<td>0.65</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>9</td>
<td>6.11</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M</td>
</tr>
<tr>
<td>cleve</td>
<td>m</td>
<td>6</td>
<td>0.02</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>8.62</td>
<td>0.07</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>13</td>
<td>60.74</td>
<td>a</td>
</tr>
<tr>
<td>hepatitis</td>
<td>m</td>
<td>6</td>
<td>0.02</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>11.42</td>
<td>0.07</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>19</td>
<td>0.26</td>
<td>a</td>
</tr>
<tr>
<td>voting</td>
<td>m</td>
<td>3</td>
<td>0.01</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>4.56</td>
<td>0.04</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>11</td>
<td>0.10</td>
<td>a</td>
</tr>
<tr>
<td>spect</td>
<td>m</td>
<td>3</td>
<td>0.02</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>7.31</td>
<td>0.13</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>20</td>
<td>0.88</td>
<td>a</td>
</tr>
</tbody>
</table>
Some of the experimental results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Minimal explanation</th>
<th>Minimum explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>size</td>
<td>SMT (s)</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>australian</td>
<td>m</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>8.79</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>14</td>
</tr>
<tr>
<td>backache</td>
<td>m</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>19.28</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>26</td>
</tr>
<tr>
<td>breast-cancer</td>
<td>m</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>5.15</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>9</td>
</tr>
<tr>
<td>cleve</td>
<td>m</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>8.62</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>13</td>
</tr>
<tr>
<td>hepatitis</td>
<td>m</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>11.42</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>19</td>
</tr>
<tr>
<td>voting</td>
<td>m</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>4.56</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>11</td>
</tr>
<tr>
<td>spect</td>
<td>m</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>7.31</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>20</td>
</tr>
</tbody>
</table>
Comparing quality to state of the art

- “Congressional Voting Records” dataset

---

Shih, A.; Choi, A.; and Darwiche, A. 2018. *A symbolic approach to explaining Bayesian network classifiers*. In IJCAI, 5103–5111
Comparing quality to state of the art

- “Congressional Voting Records” dataset
- (0 1 0 1 1 1 0 0 0 0 0 1 1 0 1) — data sample (16 features)

---

3Shih, A.; Choi, A.; and Darwiche, A. 2018. *A symbolic approach to explaining Bayesian network classifiers.* In *IJCAI*, 5103–5111
Comparing quality to state of the art

- “Congressional Voting Records” dataset
- (0 1 0 1 1 1 0 0 0 0 0 1 1 0 1) — data sample (16 features)

**Smallest size** explanations computed by:\n
- (0 1 1 0 0 0 1 1 0) — 9 literals
- (0 1 1 1 0 0 1 1 0) — 9 literals

---

³Shih, A.; Choi, A.; and Darwiche, A. 2018. *A symbolic approach to explaining Bayesian network classifiers*. In IJCAI, 5103–5111
Comparing quality to state of the art

- “Congressional Voting Records” dataset
- (0 1 0 1 1 1 0 0 0 0 0 0 1 1 0 1) — data sample (16 features)

Smallest size explanations computed by:
- (0 1 1 0 0 0 1 1 0 0) — 9 literals
- (0 1 1 1 0 0 1 1 0 0) — 9 literals

Subset-minimal explanations computed by our approach:
- (1 0 0 0 0) — 4 literals
- (1 0 0 0 0) — 3 literals
- (0 1 0 0 0 0) — 5 literals
- (0 1 0 0 0 0 1) — 5 literals

---

There are many explanations of different quality

(a) digit 1  (b) simple expl.  (c) central pixels  (d) light pixels

(a) digit 3  (b) simple expl.  (c) central pixels  (d) light pixels
Summary and future work

- *principled* approach to XAI
Summary and future work

• *principled* approach to XAI
• based on *abductive reasoning*
Summary and future work

- *principled* approach to XAI
- based on *abductive reasoning*
- applies a *reasoning engine*, e.g. SMT or MILP
Summary and future work

- *principled* approach to XAI
- based on *abductive reasoning*
- applies a *reasoning engine*, e.g. SMT or MILP
- provides *minimality guarantees*
Summary and future work

- *principled* approach to XAI
- based on *abductive reasoning*
- applies a *reasoning engine*, e.g. SMT or MILP
- provides *minimality guarantees*
- tested on ReLU-based NNs
Summary and future work

- *principled* approach to XAI
- based on *abductive reasoning*
- applies a *reasoning engine*, e.g. SMT or MILP
- provides *minimality guarantees*
- tested on ReLU-based NNs

- *other* ML models?
Summary and future work

- **principled** approach to XAI
- based on **abductive reasoning**
- applies a *reasoning engine*, e.g. SMT or MILP
- provides **minimality guarantees**
- tested on ReLU-based NNs

- **other** ML models?
- better scalability
Summary and future work

• *principled* approach to XAI
• based on *abductive reasoning*
• applies a *reasoning engine*, e.g. SMT or MILP
• provides *minimality guarantees*
• tested on ReLU-based NNs

• *other* ML models?
• better scalability
  • better *encodings*?
Summary and future work

- *principled* approach to XAI
- based on *abductive reasoning*
- applies a *reasoning engine*, e.g. SMT or MILP
- provides *minimality guarantees*
- tested on ReLU-based NNs

- *other* ML models?
- better scalability
  - better *encodings*?
  - more advanced *reasoners*?
Summary and future work

- **principled** approach to XAI
- based on **abductive reasoning**
- applies a **reasoning engine**, e.g. SMT or MILP
- provides **minimality guarantees**
- tested on ReLU-based NNs

- **other** ML models?
- better scalability
  - better **encodings**?
  - more advanced **reasoners**?
  - **abstraction refinement**?
Summary and future work

- principled approach to XAI
- based on abductive reasoning
- applies a reasoning engine, e.g. SMT or MILP
- provides minimality guarantees
- tested on ReLU-based NNs

- other ML models?
- better scalability
  - better encodings?
  - more advanced reasoners?
  - abstraction refinement?
- explanation enumeration?
Summary and future work

- **principled** approach to XAI
- based on **abductive reasoning**
- applies a **reasoning engine**, e.g. SMT or MILP
- provides **minimality guarantees**
- tested on ReLU-based NNs

- **other** ML models?
- better scalability
  - better **encodings**?
  - more advanced **reasoners**?
  - **abstraction refinement**?
- explanation **enumeration**? + **preferences**?
Questions?