

# ML pipeline exploration

Do it yourself: <https://github.com/IBM/grammar2pddl>



## Tools and links for hands-on exercise

- HTN2PDDL translation: <https://github.com/ronwalf/HTN-Translation>
- K\* planner: <https://github.com/ctpelok77/kstar>
- PlanUtils: <https://github.com/AI-Planning/planutils>
- PDDLEditor plugin for VSCode:  
<https://marketplace.visualstudio.com/items?itemName=jan-dolejsi.pddl>

# Tools and Applications

AI Planning: Theory and Practice  
IJCAI 2021

# Challenges of using AI planning in applications

- Modeling: Application domain  $\neq$  planning domain
- There are very few integrated toolchains for AI planning
- There are few libraries/frameworks that integrate AI planners
- Round-tripping is cumbersome

# Pre/post-processing

- What is the similarity/equivalence between what desired output and plans
  - Example: Are ML pipelines like plans and how? What are our ‘actions’?
- What are our inputs and is there a close-enough planning formalism?
  - Example: We have data science grammars, that look a lot like HTN specifications
- How do we map input elements to predicates and actions?
  - Example: grammar tokens = actions, grammar rules = methods
- Do we need ‘maintenance actions’ - how do we need to post-process plans?
  - Example: in the HTN to PDDL translation there are maintenance ‘method start’ and ‘method end’ actions that do not produce any pipeline tokens

# AutoAI: Automating ML Pipeline Generation



## Problem

Space of possible pipelines is huge

Humans can explore only a tiny portion of it

Humans are biased towards the pipelines they already are familiar with

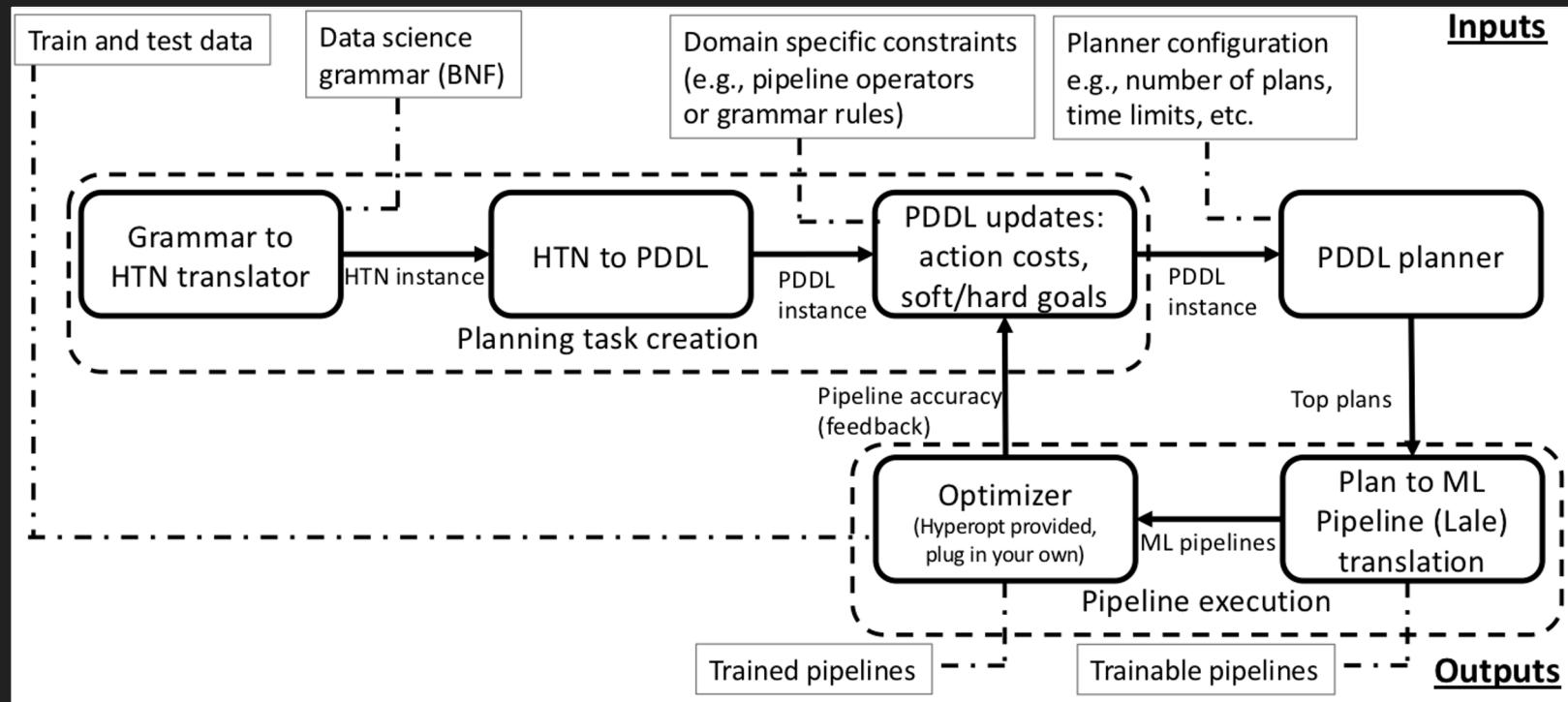
## Benefits

- Generation of pipelines of high accuracy automatically
- Reduces the need for human data scientists

## Solution

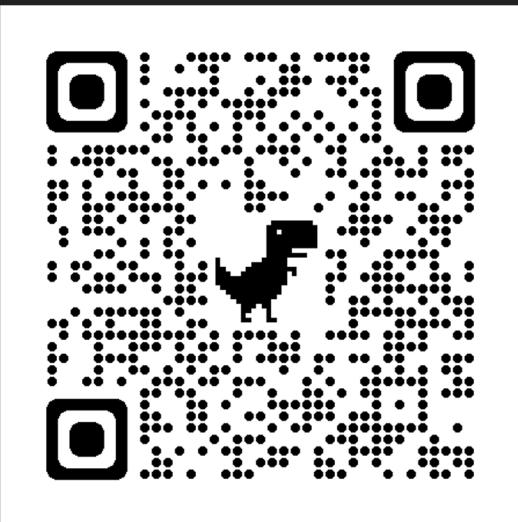
- Use regular grammars to define the possible pipeline compositions
- Translate regular grammars to HTN planning model, and then to classical planning model
- Enrich planning model with user defined constraints
- Use AI Planners to solve the planning model and translate plans to pipelines

# ML pipeline generation/exploration walkthrough



# ML pipeline exploration

Do it yourself: <https://github.com/IBM/grammar2pddl>

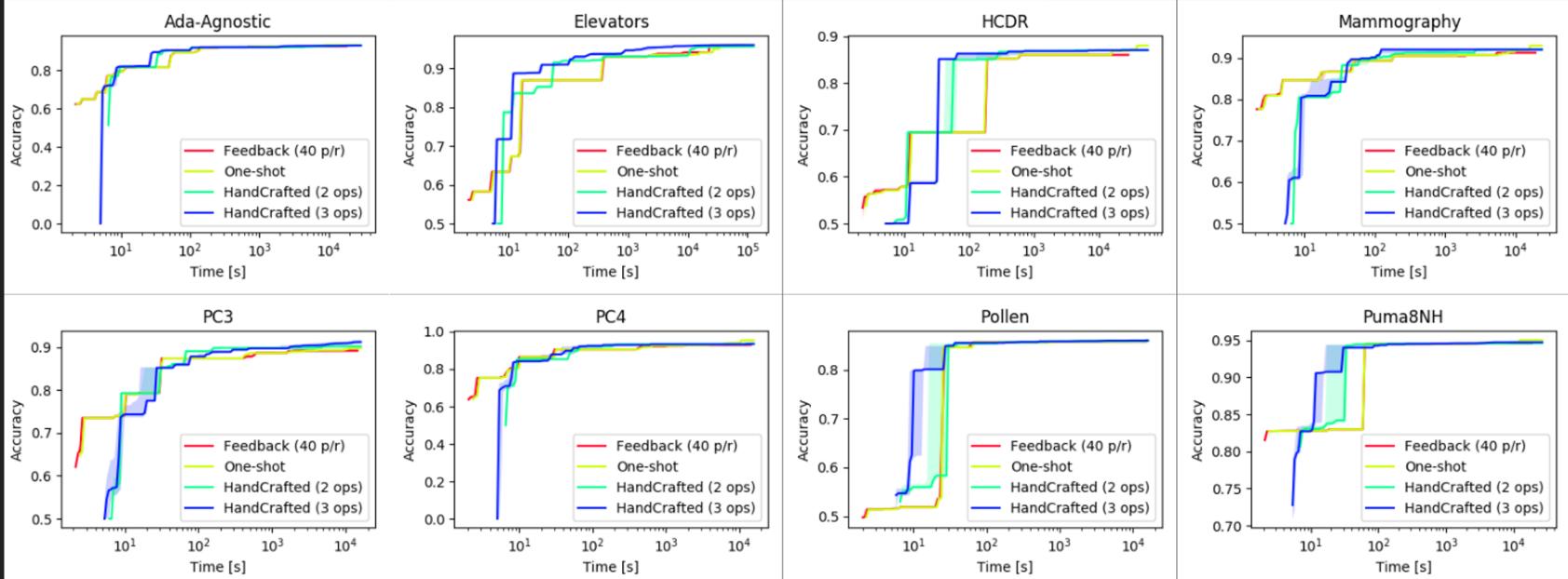


## Tools and links for hands-on exercise

- HTN2PDDL translation: <https://github.com/ronwalf/HTN-Translation>
- K\* planner: <https://github.com/ctpelok77/kstar>
- PlanUtils: <https://github.com/AI-Planning/planutils>
- PDDLEditor plugin for VSCode:  
<https://marketplace.visualstudio.com/items?itemName=jan-dolejsi.pddl>

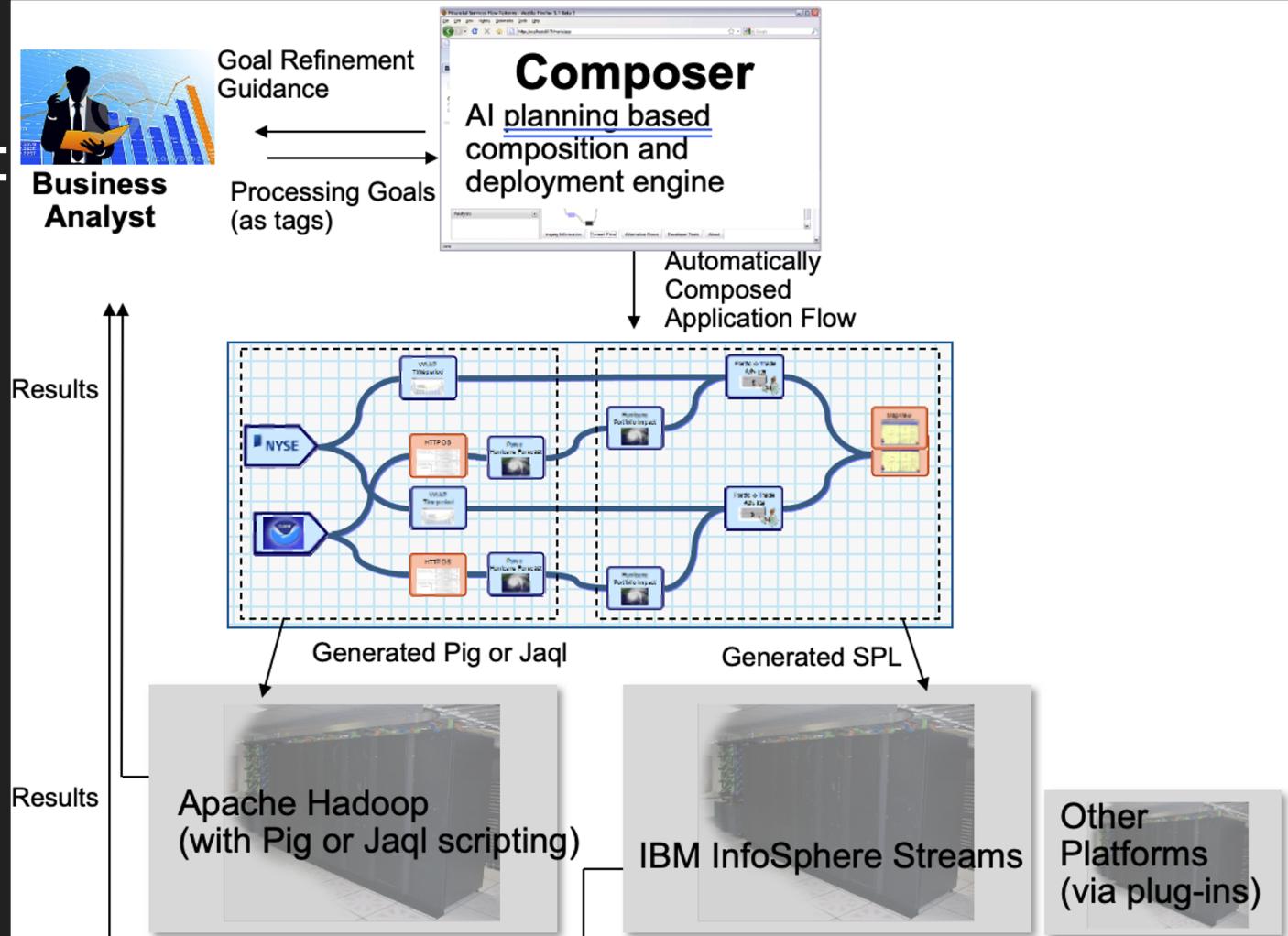
# ML pipeline exploration: matching handcrafted pipeline performance

- 27 datasets from OpenML (shown datasets with max accuracy  $\leq 95\%$ )
- Two hand-crafted baselines: manually chosen by experts chains of 2 or 3
  - W.o. feedback: one-shot 1000 plans
  - With feedback: 40 plans per iteration

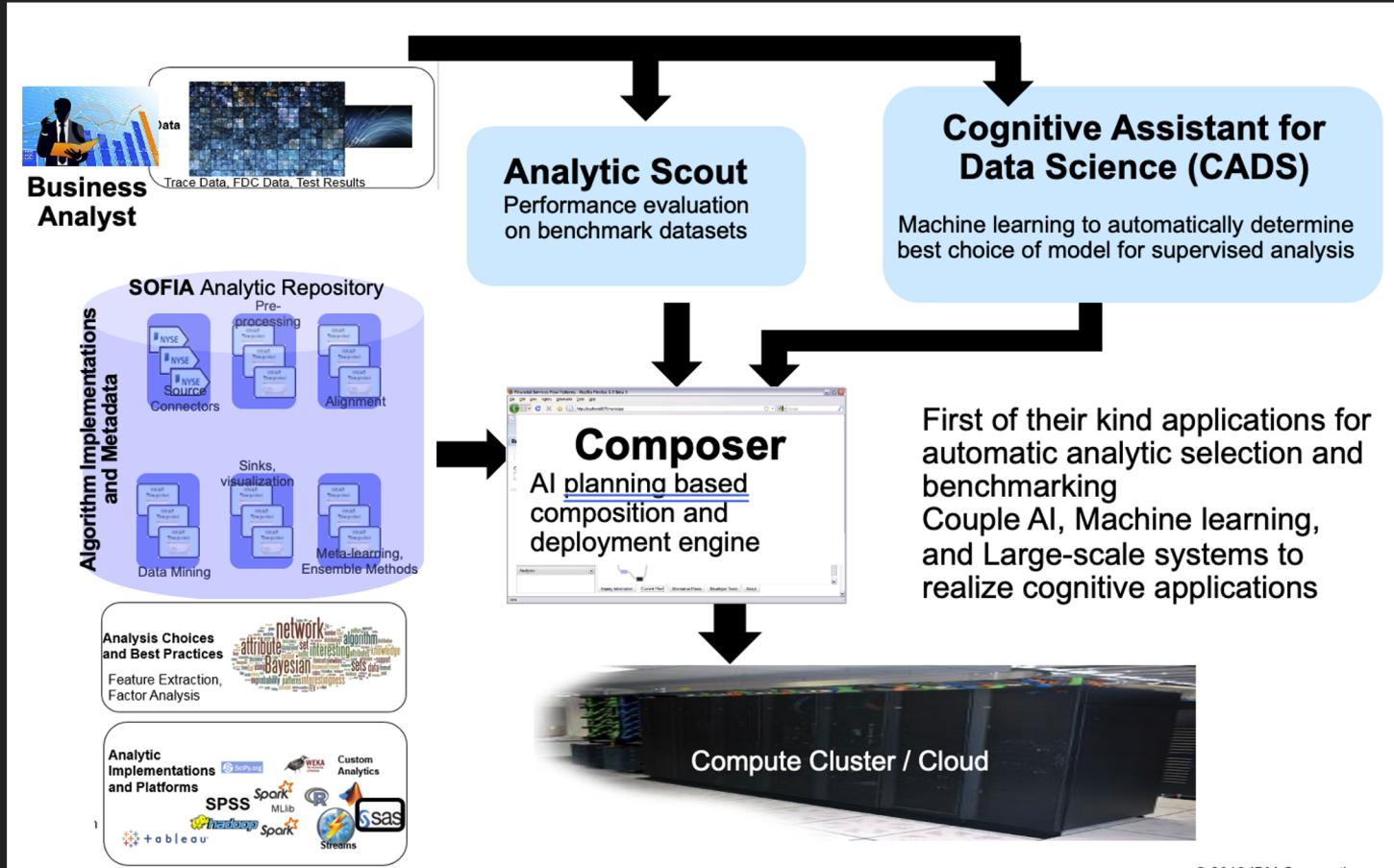


# Other Applications:

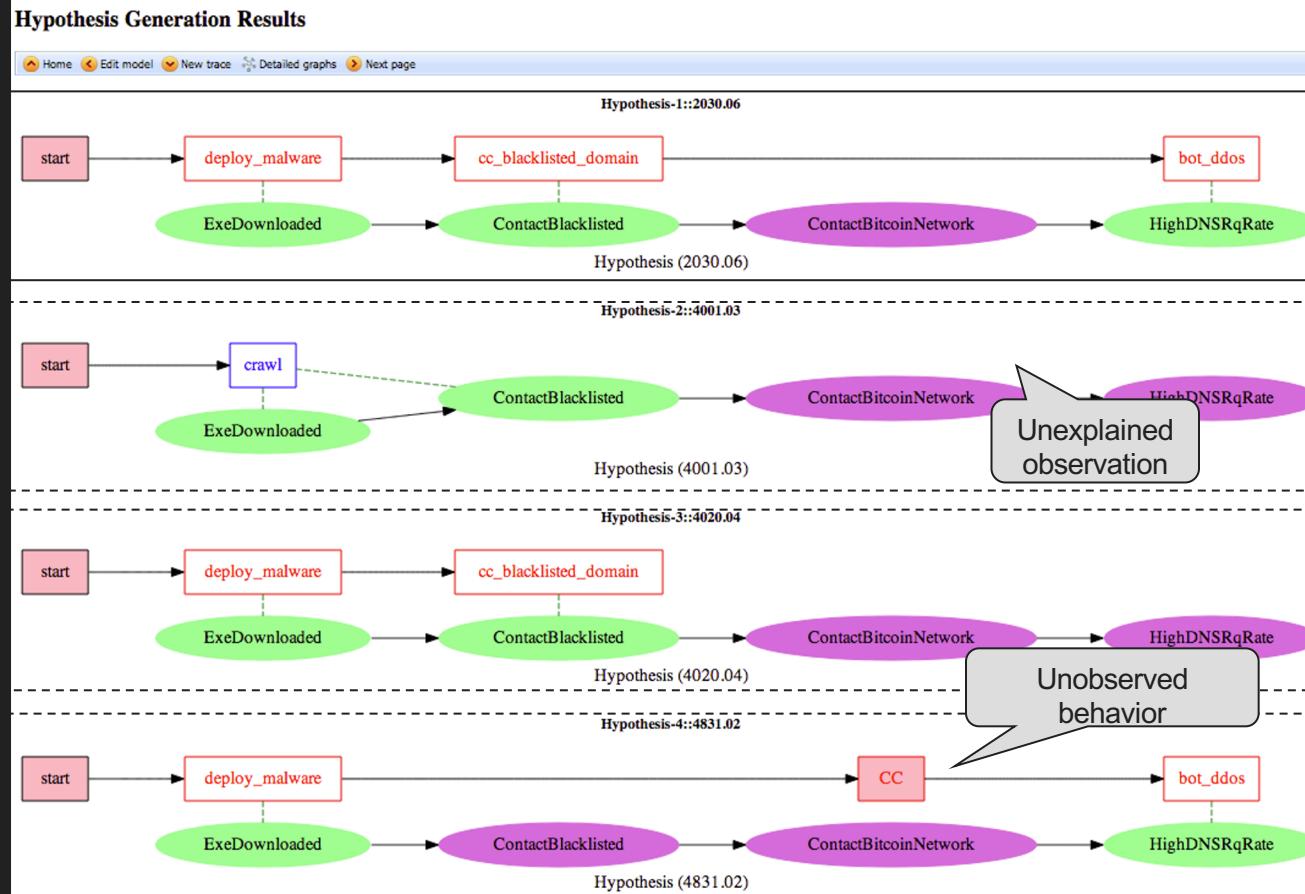
## IBM Automated Analytics Composer



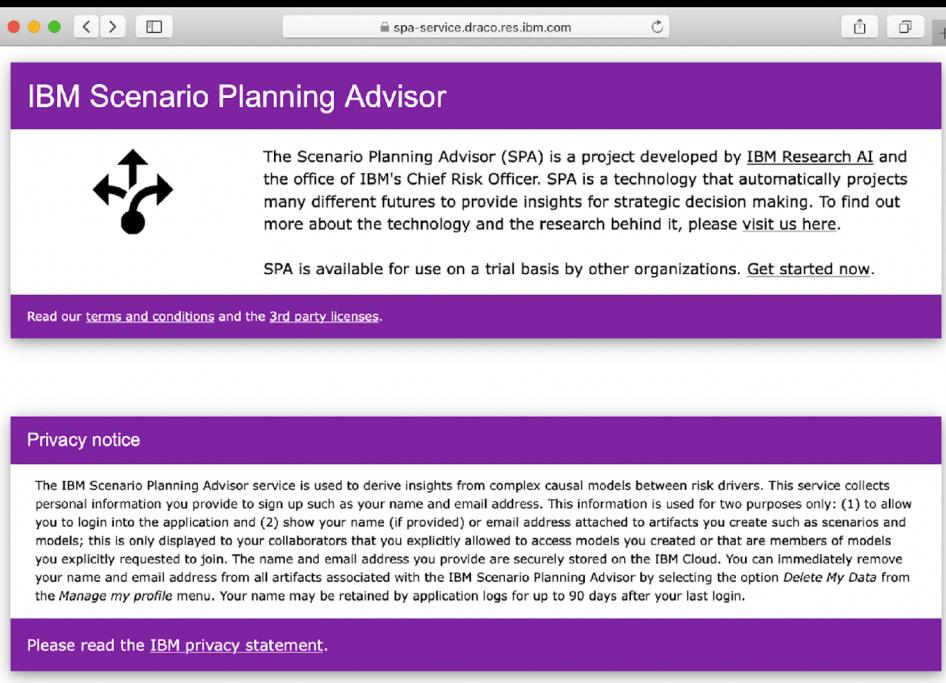
# Other Applications: DS automation via Scout and CADS



# Other Applications: Hypothesis generation in cyber



# Scenario Planning Advisor



The screenshot shows the homepage of the IBM Scenario Planning Advisor. The title "IBM Scenario Planning Advisor" is at the top. Below it is a purple header with a logo of three arrows pointing outwards from a central point. The main content area has a white background with a purple footer. The text in the main area describes SPA as a project developed by IBM Research AI and the office of IBM's Chief Risk Officer, mentioning its ability to automatically project many different futures. It also notes that SPA is available on a trial basis and provides links to terms and conditions and third-party licenses. The footer contains a "Privacy notice" section with detailed information about data collection and usage, and a link to the IBM privacy statement.

IBM Scenario Planning Advisor

The Scenario Planning Advisor (SPA) is a project developed by [IBM Research AI](#) and the office of IBM's Chief Risk Officer. SPA is a technology that automatically projects many different futures to provide insights for strategic decision making. To find out more about the technology and the research behind it, please [visit us here](#).

SPA is available for use on a trial basis by other organizations. [Get started now](#).

Read our [terms and conditions](#) and the [3rd party licenses](#).

Privacy notice

The IBM Scenario Planning Advisor service is used to derive insights from complex causal models between risk drivers. This service collects personal information you provide to sign up such as your name and email address. This information is used for two purposes only: (1) to allow you to login into the application and (2) show your name (if provided) or email address attached to artifacts you create such as scenarios and models; this is only displayed to your collaborators that you explicitly allowed to access models you created or that are members of models you explicitly requested to join. The name and email address you provide are securely stored on the IBM Cloud. You can immediately remove your name and email address from all artifacts associated with the IBM Scenario Planning Advisor by selecting the option *Delete My Data* from the *Manage my profile* menu. Your name may be retained by application logs for up to 90 days after your last login.

Please read the [IBM privacy statement](#).

## Problem

Scenario Planning for risk mitigation is a mostly manual process

Only a few scenarios can be constructed manually and explored

High impact low likelihood events are overlooked

## Benefits

- Reduction in time for building scenarios from months to hours
- Exploration of orders of magnitude more scenarios than possible if built manually

## Solution

- Exploit NLU techniques to semi-automatically construct scenario planning models
- Automatically explore the space of possible scenarios with an AI Planner
- Choose scenarios of high relevance to a client at a particular time