How to assure the software quali
for Artificial Intelligence?

~ Beyond the artificial intelligence Era and its accountability ~

July 7th, 2017

Nobuhiro Hosokawa (CARVIN@jp.ibm.com)

Security and Service,
IBM Research-Tokyo.

Today's topic : Quality Assurance for AI and Machine Learning

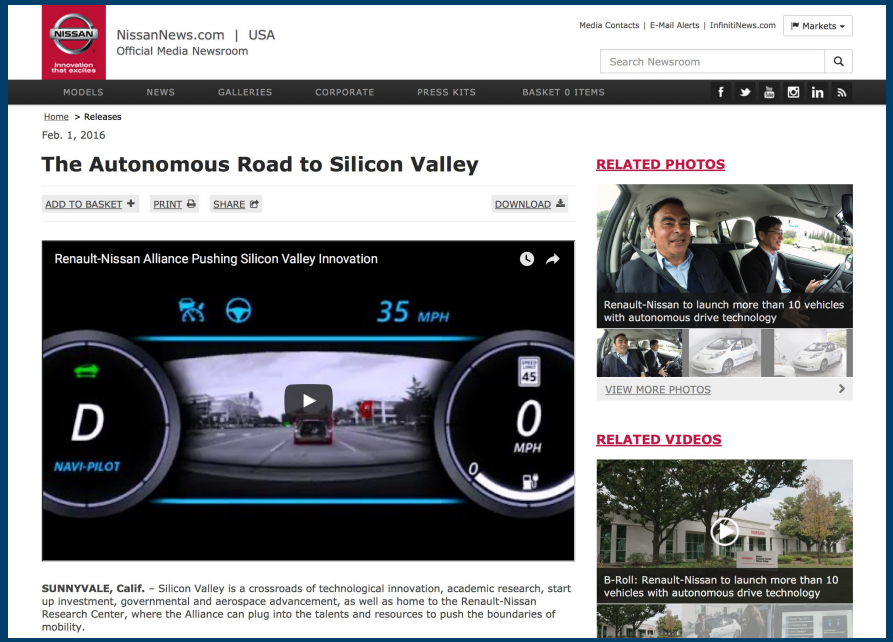
- **[Problem]** Machine Learning(ML) solutions sometimes are hard to be adopted, since the accountable explanations of outputs from machine learning modules are difficult.
- **[What are the attacks?]**
  - In the Autonomous Driving, Safety driving system with Artificial Intelligence(AI) technologies (ex: Auto braking system with Visual Recognition, Auto Lane Change, Across the Intersection with Machine learning) market is rapidly growing. "Auto driving Car" is nearly reached to "LEVEL 5".
  - There is no process to validate and to verify (=proof) the correctness of decision made by ML
    - Need to provide the test, review technology for Cognitive system
    - Establish the process to proof the Traceability, Accountability, and so on.
- **[What are the attacks consequences?]**
  - Commoditization of Artificial Intelligence system.
  - More widespread use of safer and reliable AI system.
- **[Where is the data coming from? How to control it?]**
  - Private data: Each of auto maker have the personal (=private) driving data that ware collected through connected car system.
  - Public Data: there is no public data.

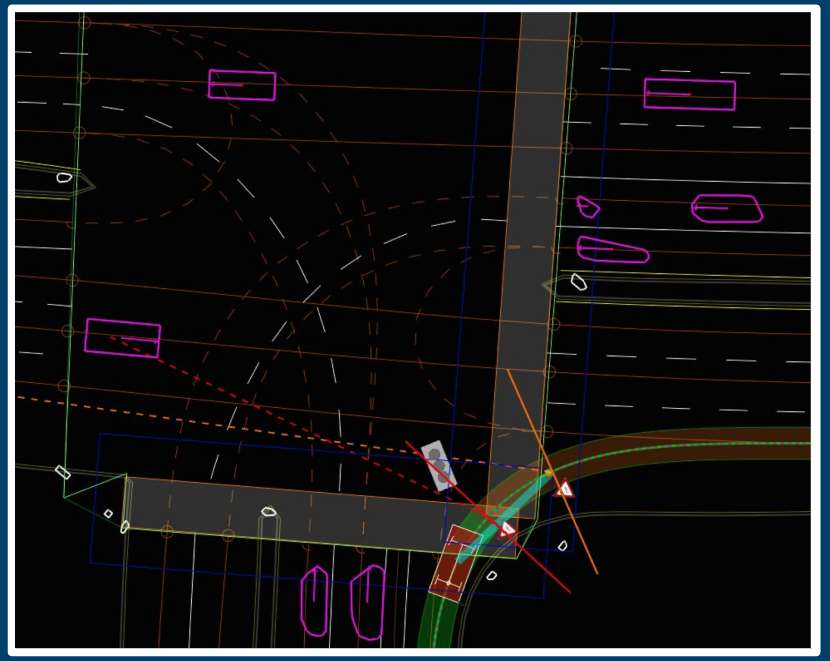
- **[What's Happening Outside?]**
  - Political topic: In Oct.2015, United Nations Parliamentary Congress started to discuss about the risk/threat of Super Intelligence, including chemistry, biology, Radioactive substances, Nuclear (CBRN) weapons, Autonomous Driving. Discussion are concluded as In short term, the impact of artificial intelligence will depends on "who should control it", but in long term, "This is issue about 'Is Artificial Intelligence controllable or not ?' "
  - Market Topic: Google and Nissan publish their effort to establish the autonomous driving technologies from 2014.

Today's topic : Quality Assurance for AI and Machine Learning

- **[Can we quantify the risk?]**
  - **Problem that can not be "verified" of the plasticity of the neural network.**
    (= Can not be pursued later if artificial intelligence has concluded to determine)
    - Previous research in the field does not exist.
  - Security and Privacy : Driving data for autonomous driving might be categorized as the privacy data.
  - Regal Regulations: ex) ISO26262…
- **[Example from the papers]**
  - Verification and validation of neural networks: a sampling of research in progress
    https://people.cs.umass.edu/~btaylor/publications/PSI000008.pdf
  - Probabilistic Model Checking
    http://www.prismmodelchecker.org/bibitem.php?key=KNP07a
  - Using probabilistic model checking in systems biology
    http://dl.acm.org/citation.cfm?id=1364651
  - EXE: Automatically Generating Inputs of Death
    http://dl.acm.org/citation.cfm?id=1455522&CFID=755989418&CFTOKEN=93994630
  - Towards Automatic Discovery of Deviations in Binary Implementations with Applications to Error Detection and Fingerprint Generation
    https://www.usenix.org/legacy/events/sec07/tech/brumley/brumley_html/paper.html

Ex1) Scenario : Self-Driving Car (Autonomous Mode)





Issues need to be discussed

Same issues are existing in following areas

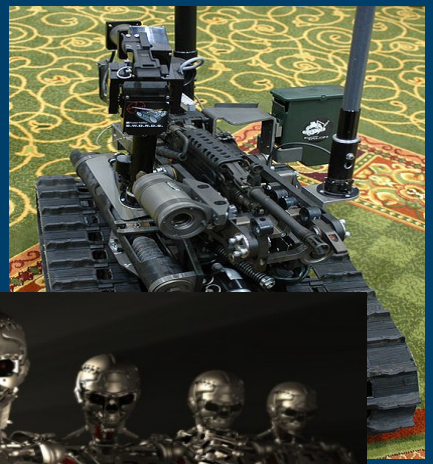
- AI Weapon

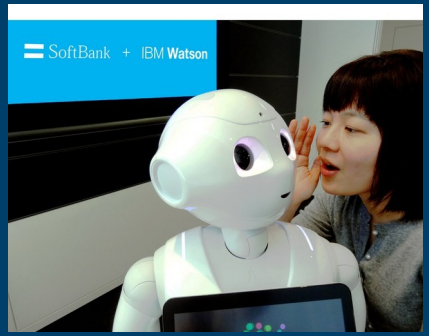
  as 3rd generation of weapon. Discussed in US Government. Aero & Defence area
- AI for Healthcare

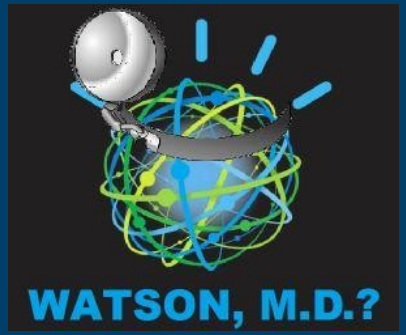
  Lots of legal regulation : ISO 13485, 14971…

  Data privacy : Huge number of medical records.
- Robot + AI

  Robotics. Especially, testing for robotics. As in Fintech/FSS area.

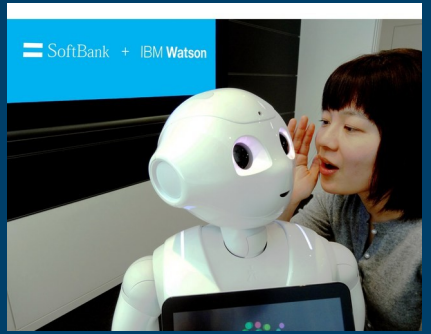






Ex2) Characteristics of Cognitive robot system (such as Pepper + Watson)

- Cognitive robot systems are basically constructed as *Reactive Systems*
  - Robot system needs to interact with a lot of sensing and action modules like STT(speech-to-text), TTS(text-to-speech), human detection face recognition contact recognition, robot's hand , face and body control
- Reactive system
  - https://thinkit.co.jp/article/9185
  - https://en.wikipedia.org/wiki/Reactive_system
  - https://en.wikipedia.org/wiki/Reactive_programming



  - Main parts of system behave reactively (not a part of the system)
  - Flow-based processing is in rather local parts.

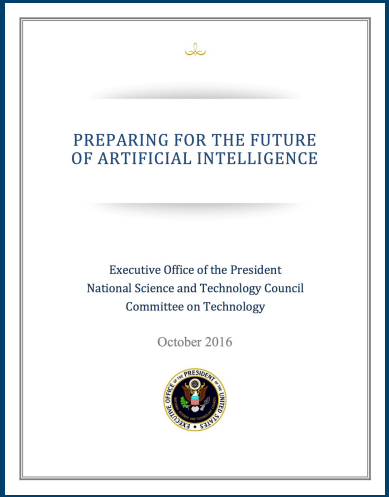
  - System behaviors are sensitively depend on input timing of events
  - Each event is raised in continuous time line (not discrete time line) and need to be reactively handled.

  - Non blocking message/Event passing implements module communications
  - Shared state like shared memory are not favored
    - It declines the reactivity because of resource conflicts such as deadlock and livelock
    - It downgrades resilience because of high module coupling degree .
      - A problem in a module tends to cause problems in other modules

Ref) Problems in Testing of Cognitive Robot system

- Events in continuous time line
  - It is difficult to comprehend and model target application as path- or flow- model in traditional way
- No Bug Reproducibility
- Many bugs seems to be no reproducible because processing sensitively depends on timing of events in real time
- Difficult to comprehend application state
  - Reactive system incline to use no shared states, so it is difficult to comprehend application current state.
- No mature framework for testing Event-Condition-Action rules
  - Theoretically such tool may be exists but no practical mature tools
  - No debugger for such ECA rule reactive system (like one-step execution debugger for imperative programming model)
- Depending on low quality hardware module (that may go into fault in many times)
  - Difficult to decide whether each bug is caused from hardware or software
  - For example, each pepper working for pepper is replaced 5 times in a month.
- Real world gives Input events(like speech event) and receives output (like raise hand action)
  - Correct behavior is difficult to be comprehensively described
  - Many test scenarios that cannot be handled
- No test completion criteria
  - Difficult to measure test completion rate and criteria (Traditionally path-coverage or branch-coverage)

Oct. 2016 : “Preparing for the future of AI” from USA



PREPARING FOR THE FUTURE
OF ARTIFICIAL INTELLIGENCE

Executive Office of the President
National Science and Technology Council
Committee on Technology

October 2016

Distributed material

Proposal of Discussion toward Formulation of AI R&D Guideline

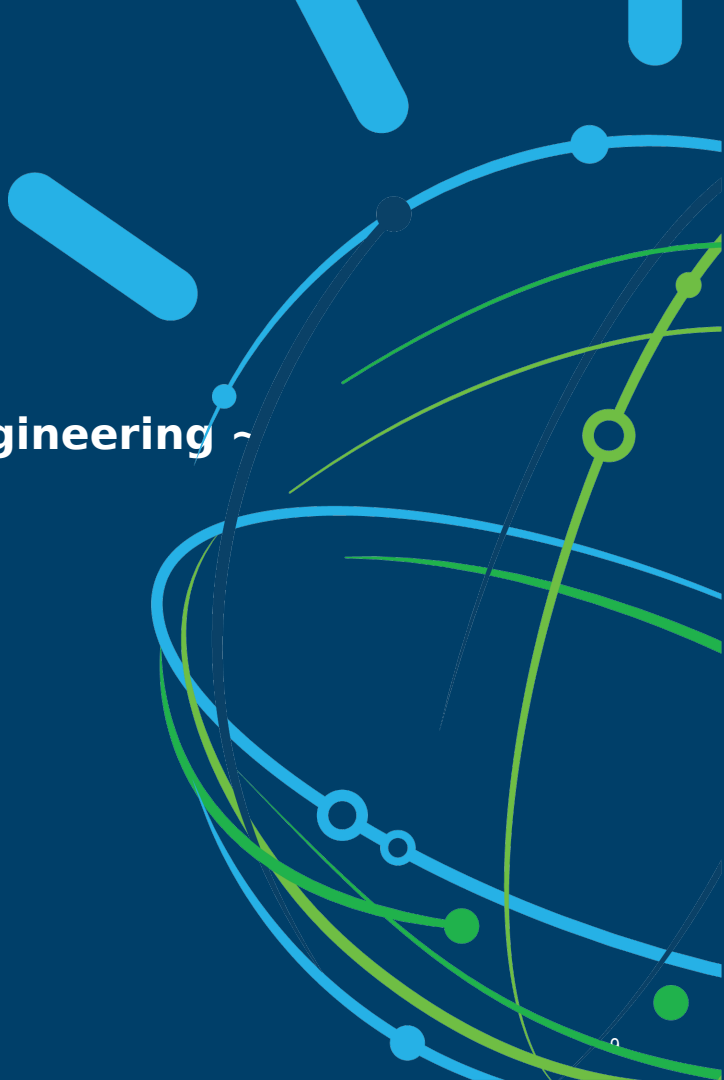
Referring OECD guidelines governing privacy, security, and so on, **it is necessary to begin discussions and considerations toward formulating an international guideline consisting of principles governing R&D of AI to be networked** (“**AI R&D Guideline**”) as framework taken into account of in R&D of AI to be networked.

Proposed Principles in “AI R&D Guideline”

1. **Principle of Transparency**
   Ensuring the abilities to explain and verify the behaviors of the AI network system
2. **Principle of User Assistance**
   Giving consideration so that the AI network system can assist users and appropriately provide users with opportunities to make choices
3. **Principle of Controllability**
   Ensuring controllability of the AI network system by humans
4. **Principle of Security**
   Ensuring the robustness and dependability of the AI network system
5. **Principle of Safety**
   Giving consideration so that the AI network system will not cause danger to the lives/bodies of users and third parties
6. **Principle of Privacy**
   Giving consideration so that the AI network system will not infringe the privacy of users and third parties
7. **Principle of Ethics**
   Respecting human dignity and individuals’ autonomy in conducting research and development of AI to be networked
8. **Principle of Accountability**
   Accomplishing accountability to related stakeholders such as users by researchers/developers of AI to be networked

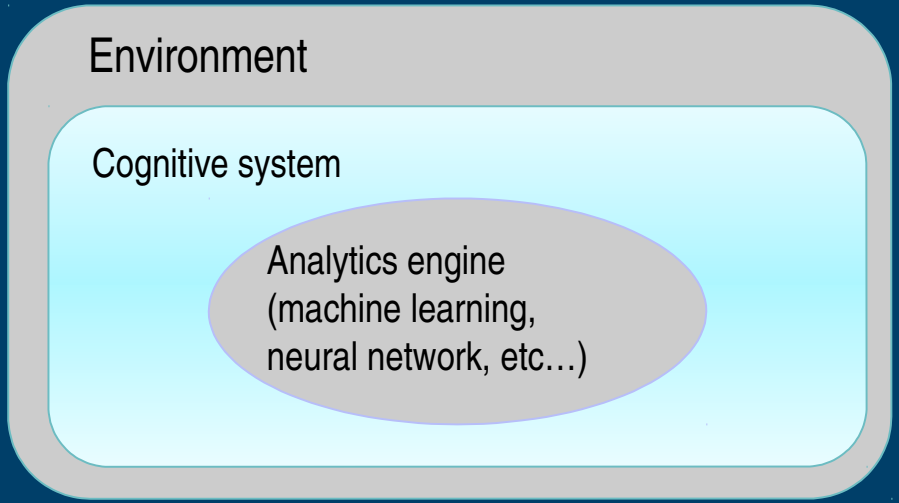
Solutions :

~ Targeting, Logical Reasoning and Defect Engineering ~

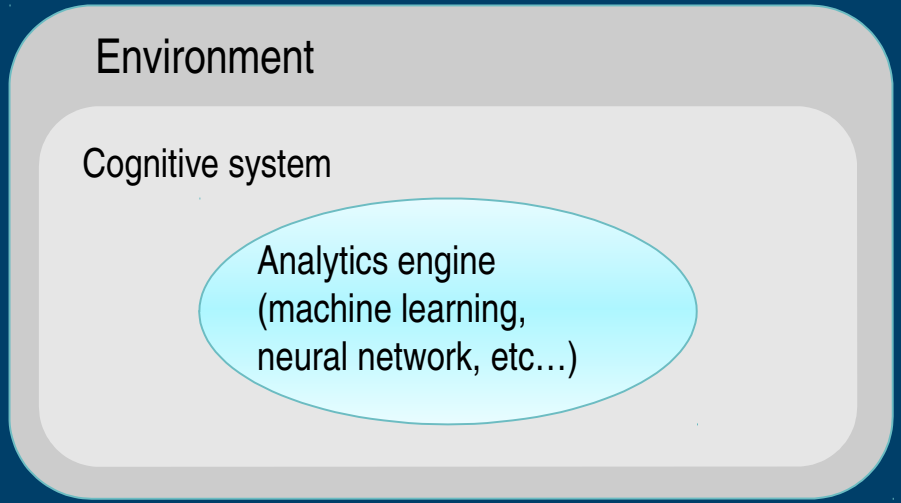


Approach 1) Which is the target? : Targeting to assure

System boundaries and environments for V&V



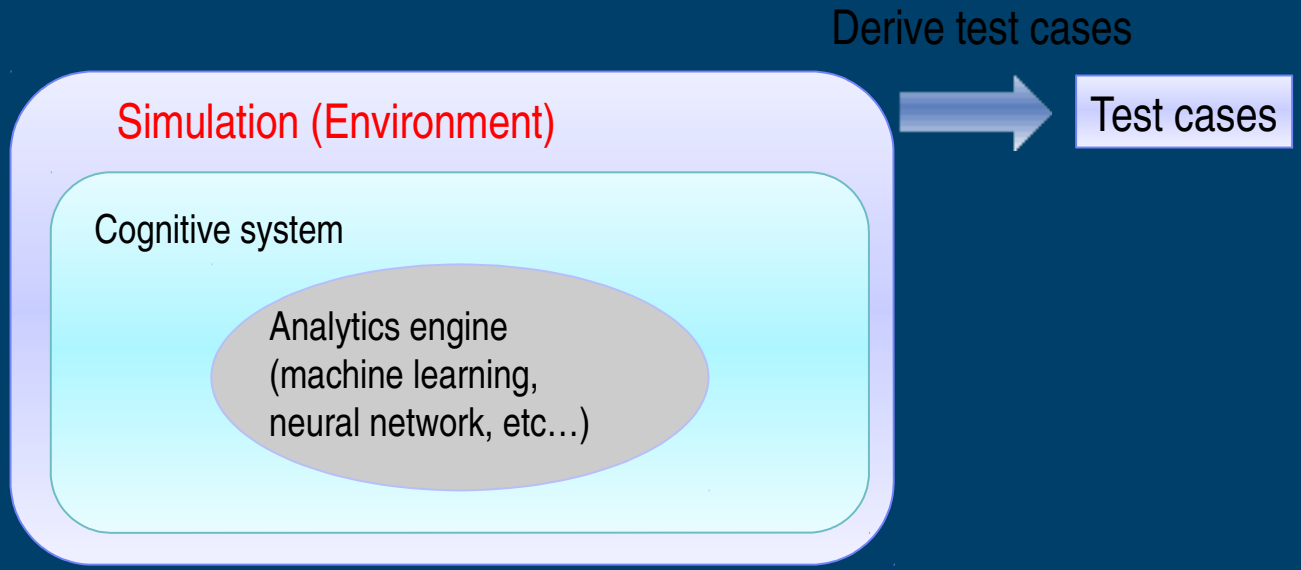
Verify & validate the cognitive systems with the models of the environment and the analytics engine.



Verify & validate the cognitive engine.

Note: the cognitive system includes a component that encodes the physical and/or logical models as vectors.

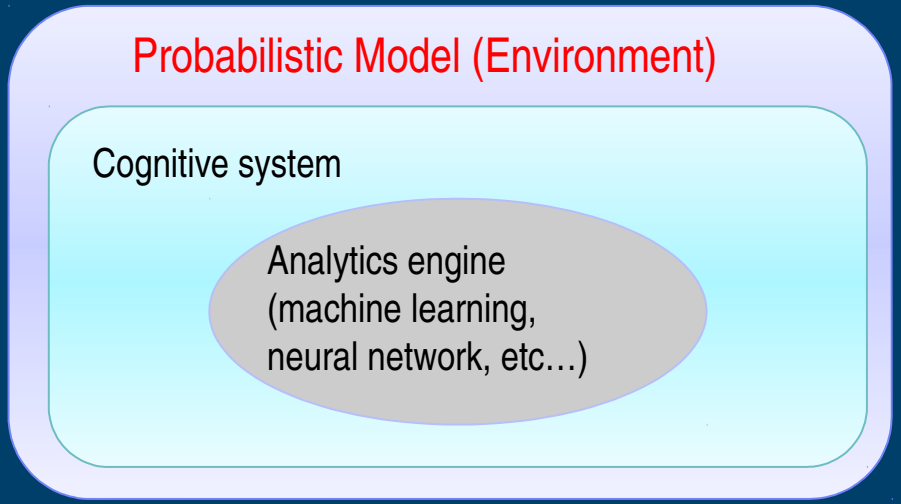
Current Approach for Software Testing



Problems:

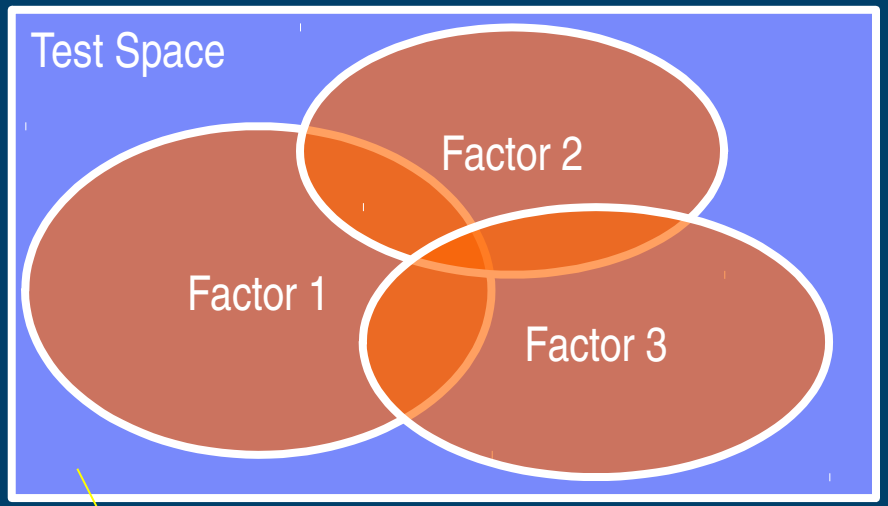
- Difficult to explain how accurately the simulation represents real-world phenomenon.
- Difficult to model the real world, and to explain the coverage of test cases derived from the simulation model.

Next Generation of V&V for Cognitive Systems



Key Technologies (to complement the simulation-based approach):
- Stochastic model checking (for checking % of correctness of the system)
- Probabilistic reasoning / abduction (for explaining the environmental scope of the system)

Approach 2: Too much test cases -> Contradiction approach



1) Test Planning
- Approach : “Proof by Contradiction”
- Test “Complement” first
  - “Evil Case” = Worst Scenario define
  - Domain specific knowledge needed

2) CTD : Combinatorial Test Design
  - How quality can be kept with reducing the test case?
  - Physical / Logical Test Gap.

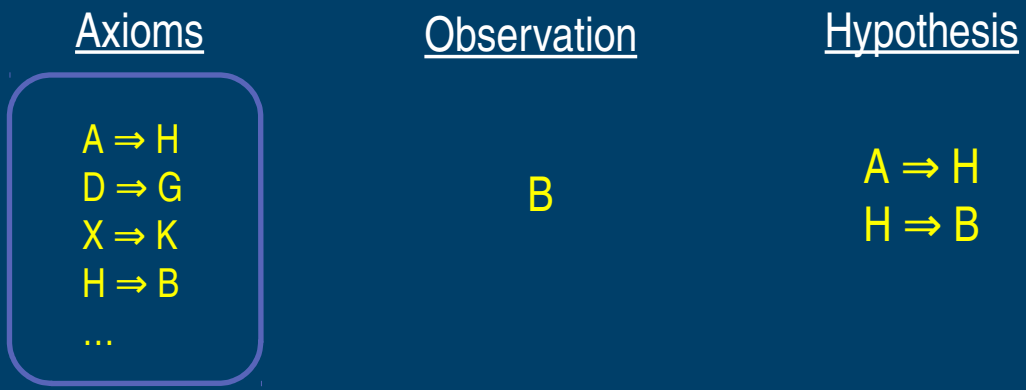
Can we automatically design the test case with Cognitive / machine learning system?

Approach 3 : Logical Reasoning

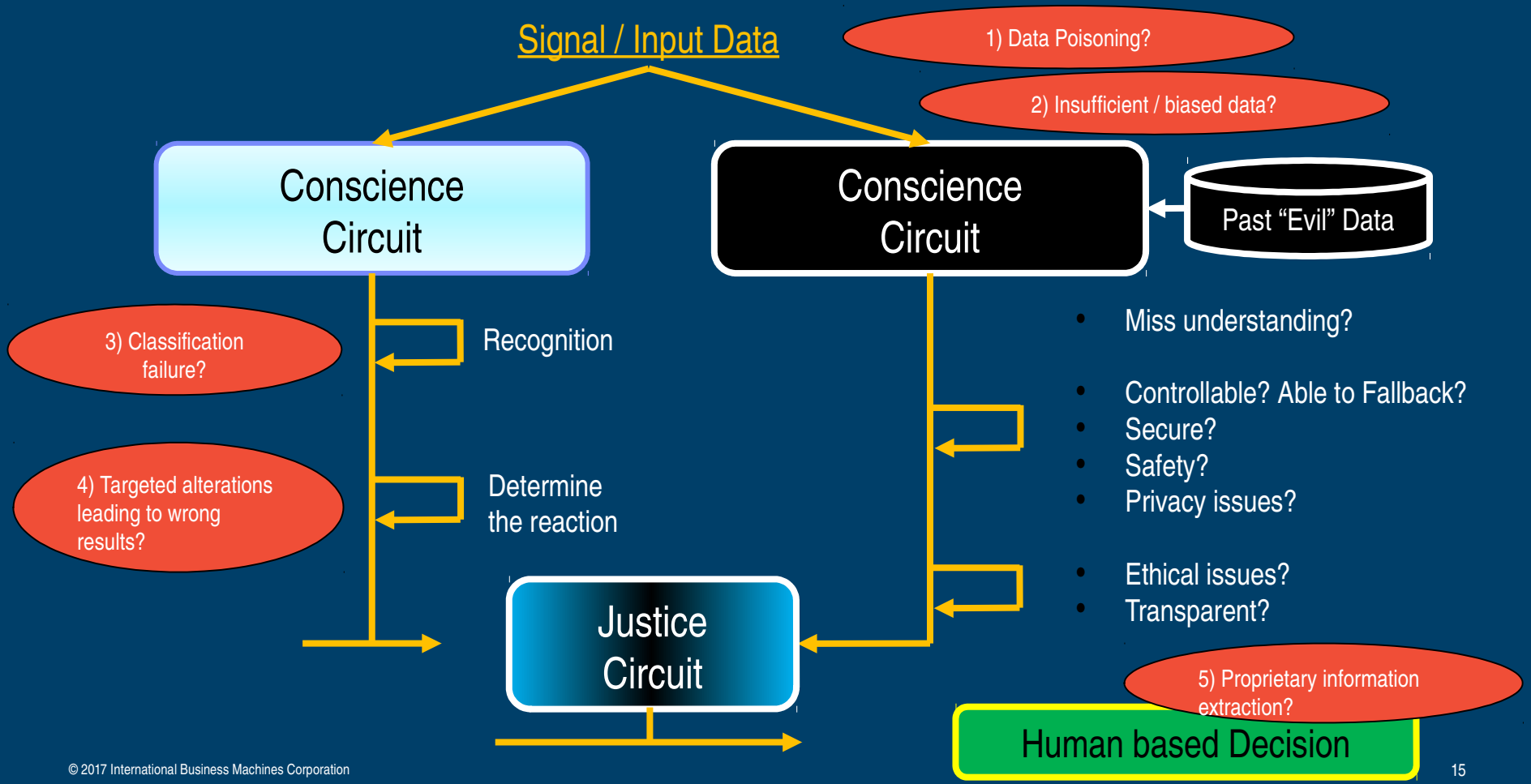
Deduction $A \Rightarrow B$

Induction $A1 \Rightarrow B, A2 \Rightarrow B \ldots$ infers all $Ai \Rightarrow B$

Abduction Given B, find a sequence of hypotheses which well describes B.



Conscience Circuit with Evil Circuit



How to create Evil Data? :Defect Engineering using NLP

IBM

Incident reports

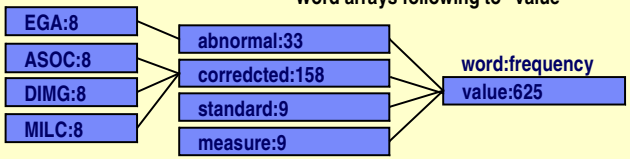
XXXXX

A) Create “Defect Model”

1) Classification : Defect Keywords Tree

□□ (customer-specific Defect Classification Tree)

Word arrays following to “value”



Validating defect keyword tree with the domain experts

2) Convert to “Keyword Hash Map”

Function	Location	Value	Timing
function AA	XX Stage	corrected value	Job restart
Function BB	YY Unit	Standard value	F condition

Abstract the incident knowledge to use all quality improvement movements

Term Extraction

Contents Analytics

Ex) Extract words related to “value” such as “EGA abnormal value”, “ASOC corrected value”
(called “PrefixSpan” technique)

B) use “Defect Model

Watson

Feature #4 Modify YY

Fix the incident in Func BB=> Adjusts “ZZ”

Feature #3

Word ”YY” must be included. But does not exist in Specification!!

Feature #2 Fix the incident in Func AA =>Implement “XX”

Initial Future

Func “AA” □□ 100 times
Func “BB” □□ 20 times
Func “CC” □□ 10 times

In the first feat… the pair… has map : AA-XX was …

Ex) Automatically review the specification with WCA and harvested
Keyword Hash Map

Wrap up

- There is no complete method to assure the AI Quality, right now.

- But, some techniques are useful for assurance.
  - Targeting approach : specify the scope to assure the quality.
  - Logical Reasoning : Abduction to find the reason why the result was come.

- Conscience Circuit & Evil Circuit : Justice with pessimistic thinking
  - Several perspective of principles for AI quality assurance
  - Robustness for the AI Quality : against “Data poisoning”, “Insufficient data”, “Biased data”, “Classification Failure” …

- Defect Engineering : Need to collect the defects past experienced.
  - To collect the defect with Natural Language Processing, Data Analytics Patterns, features, attributes.…
  - Evil data : knowledge about defects, failures are the key to success the AI area.

The future is always created through transformation
of ideas and learning of mistakes.

The future of AI that we hope will be made with our hands.

July 2017, Nobu