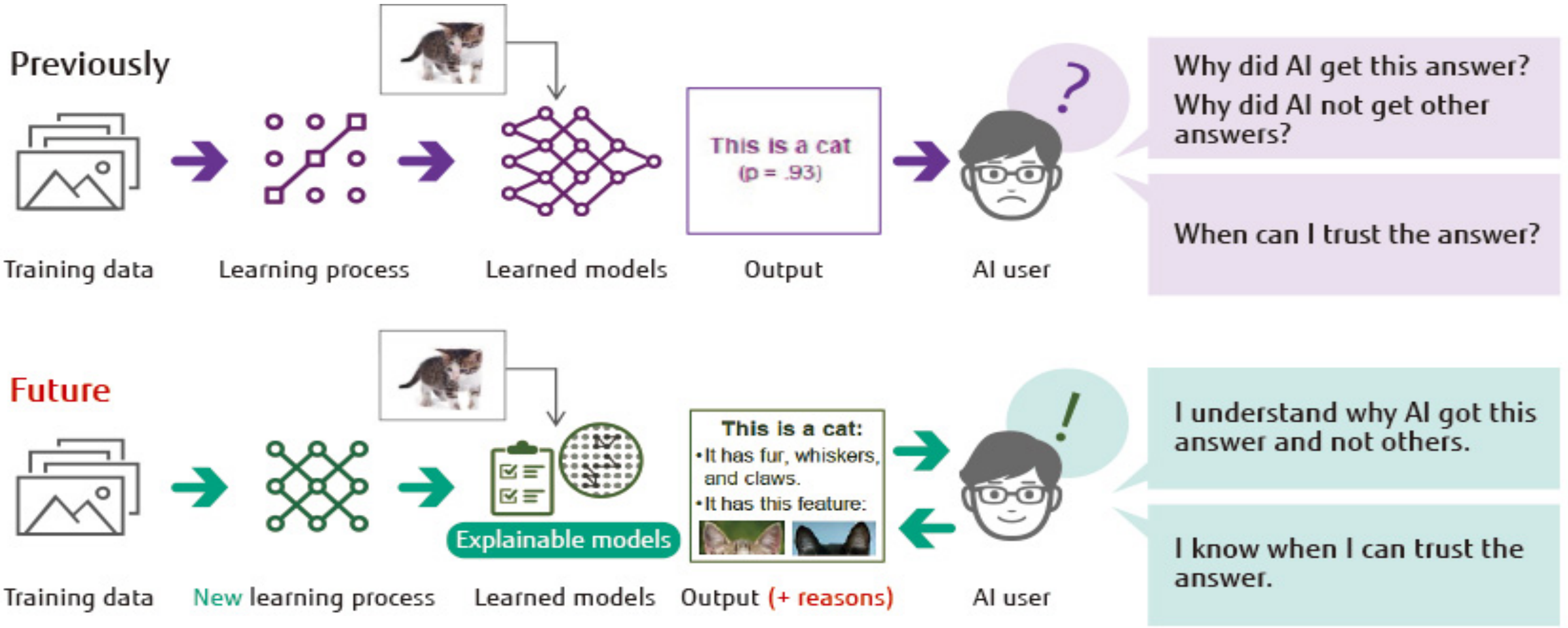
Explainable AI: explain what to whom?

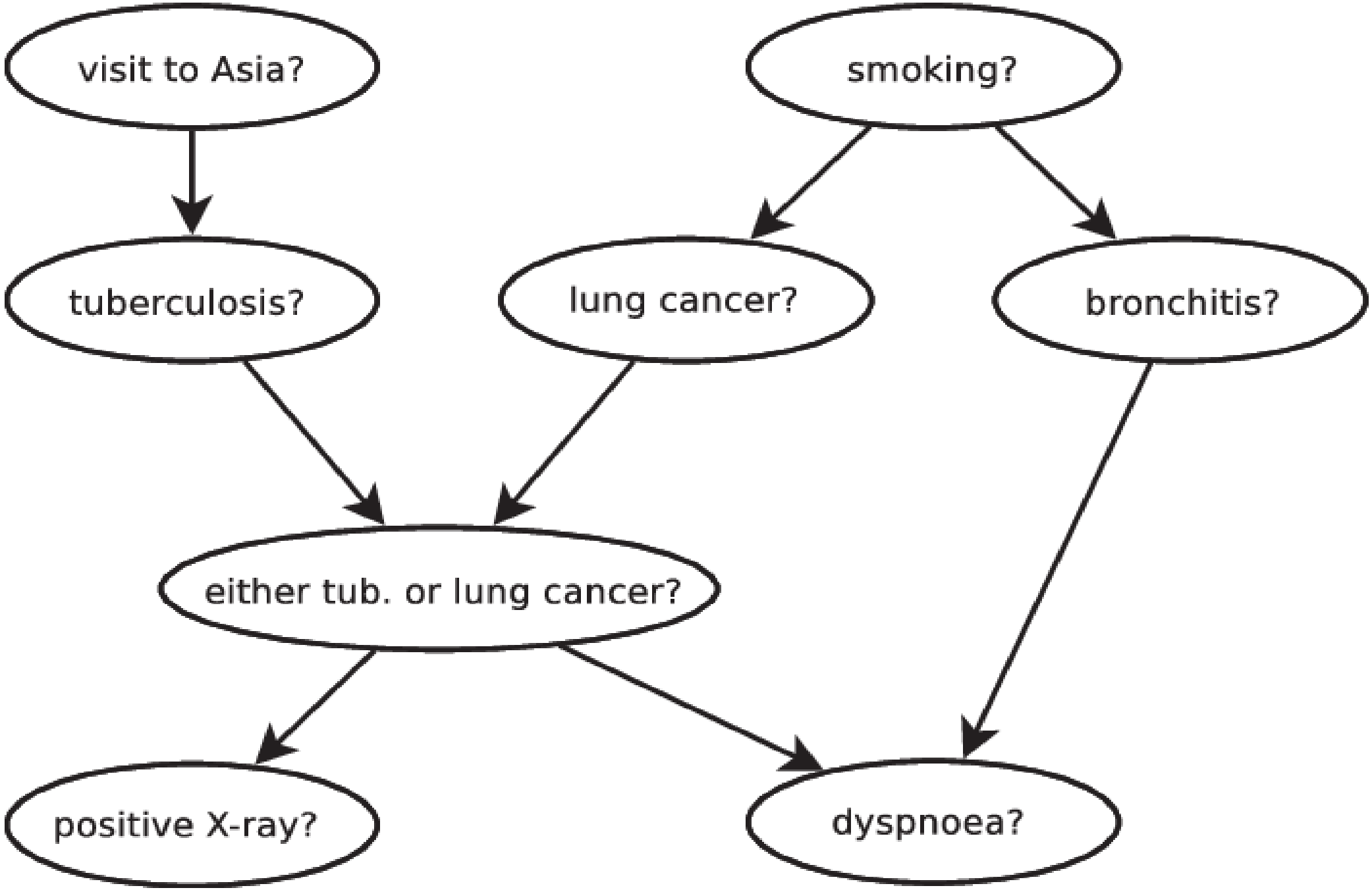
Silja Renooij

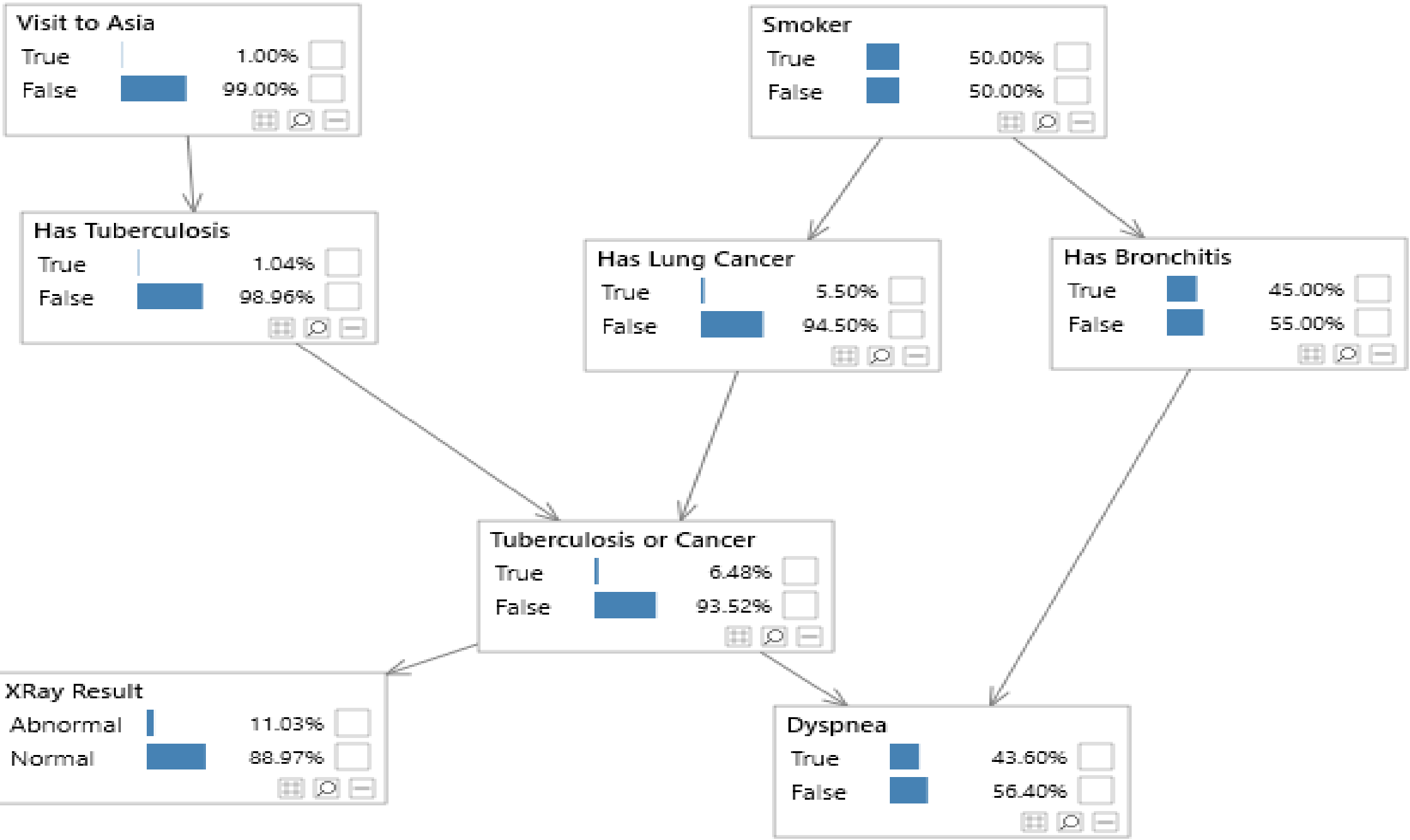
The goal of Explainable AI

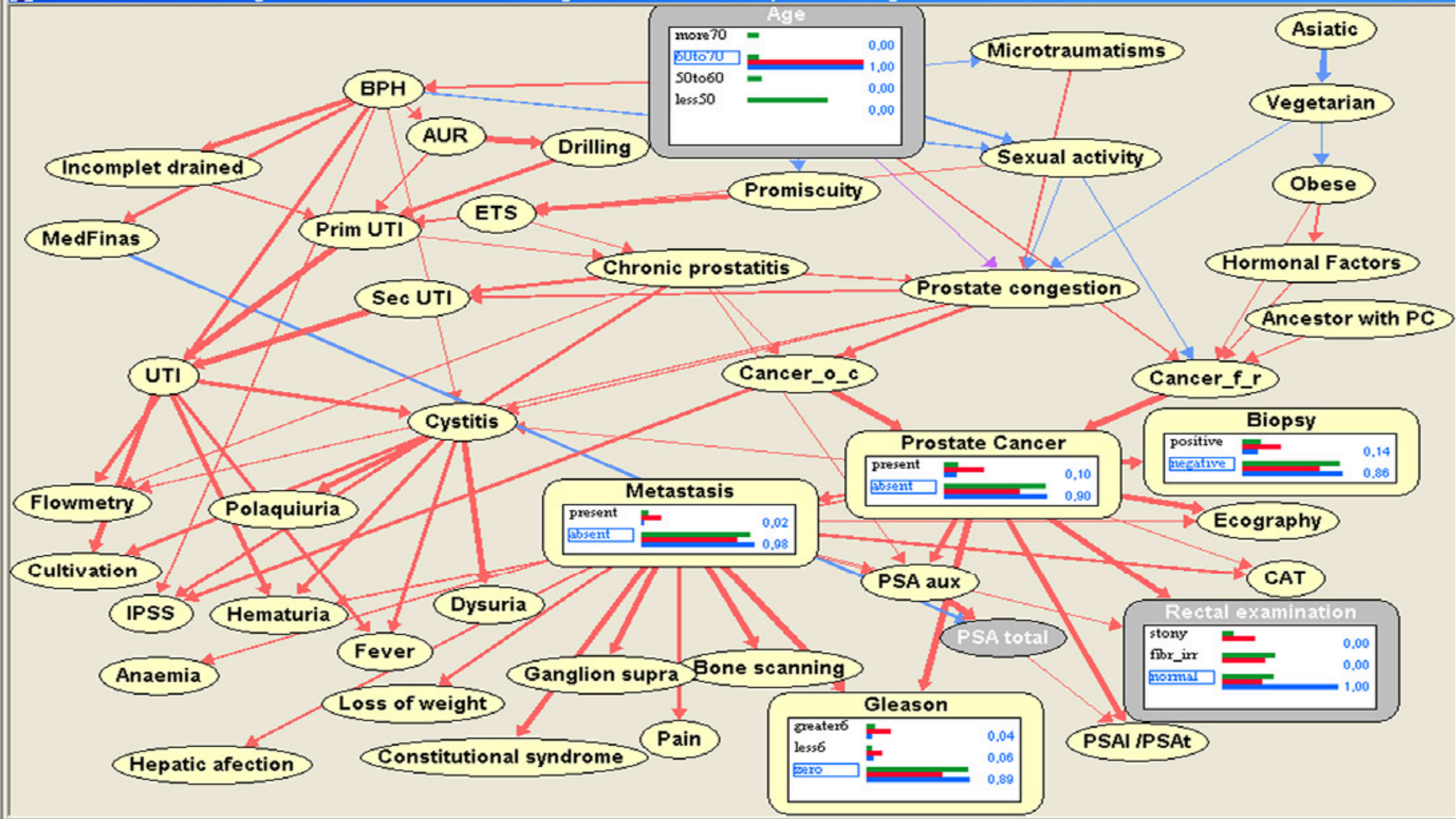


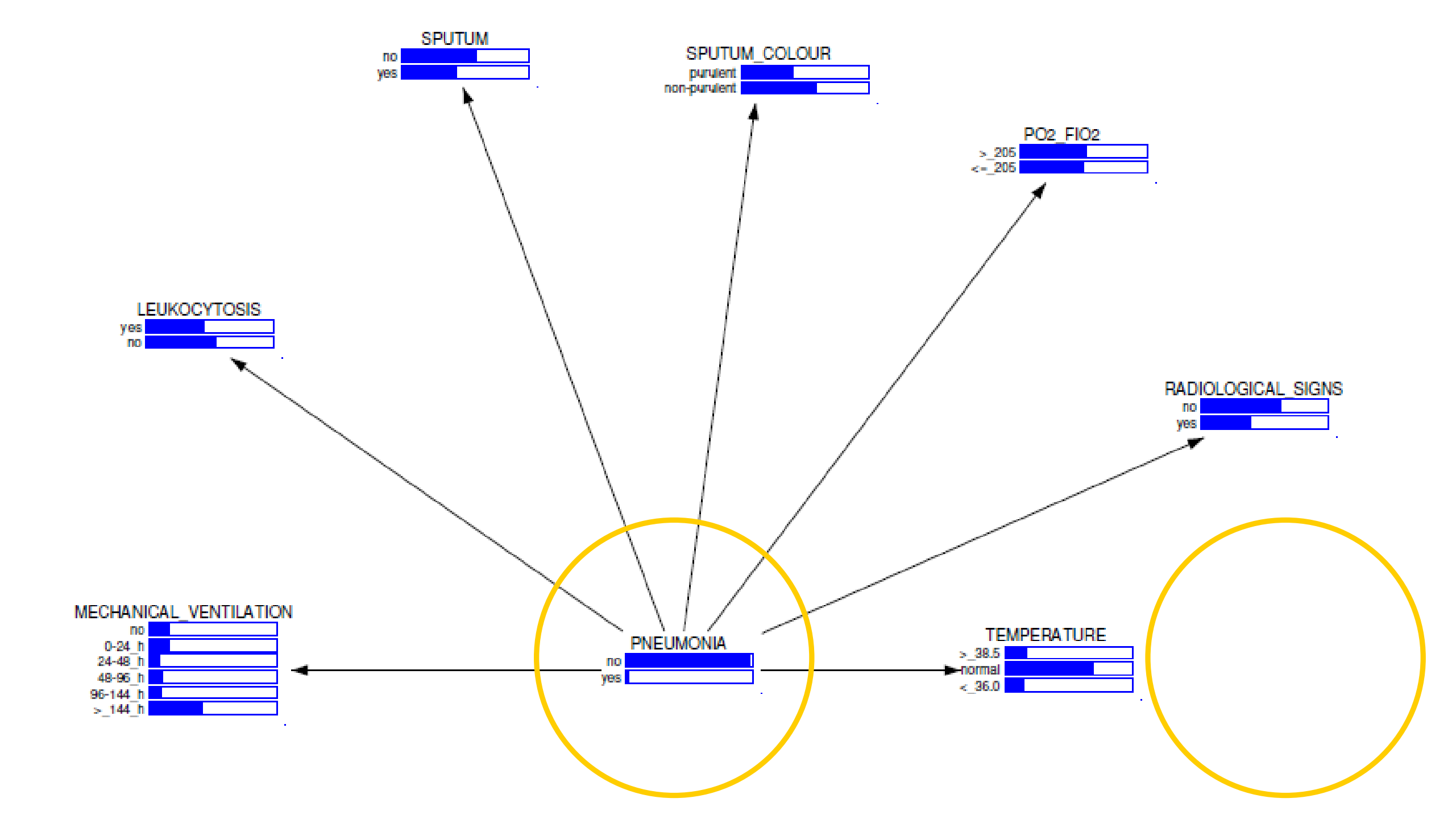
Wikipedia:

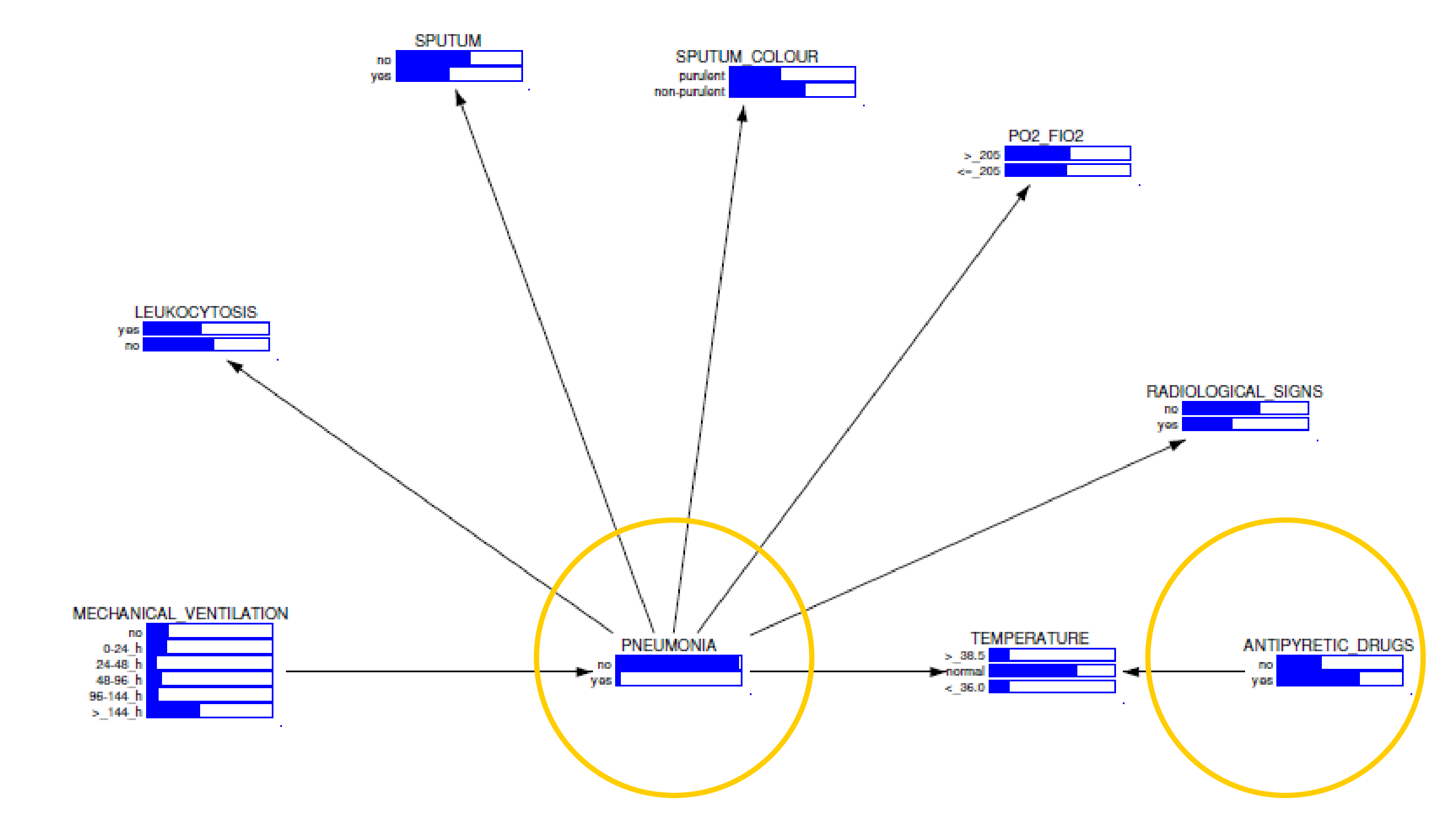
Explainable AI (XAI) refers to methods and techniques in the application of artificial intelligence technology (AI) such that the results of the solution can be understood by human experts.


visit to Asia?
smoking?
tuberculosis?
lung cancer?
bronchitis?
either tub. or lung cancer?
positive X-ray?
dyspnoea?


Visit to Asia
True
1.00%
False
99.00%
Smoker
True
50.00%
False
50.00%
Has Tuberculosis
True
1.04%
False
98.96%
Has Lung Cancer
True
5.50%
False
94.50%
Has Bronchitis
True
45.00%
False
55.00%
Tuberculosis or Cancer
True
6.48%
False
93.52%
XRay Result
Abnormal
11.03%
Normal
88.97%
Dyspnea
True
43.60%
False
56.40%


Age
more70
50to70
50to60
less50
0,00
1,00
0,00
0,00
Asiatic
Microtraumatisms
Vegetarian
BPH
AUR
Drilling
Incomplet drained
Sexual activity
Obese
Promiscuity
ETS
Prim UTI
MedFinas
Chronic prostatitis
Hormonal Factors
Prostate congestion
Sec UTI
Ancestor with PC
UTI
Cancer_o_c
Cancer_f_r
Cystitis
Biopsy
positive
negative
0,14
0,86
Prostate Cancer
present
absent
0,10
0,90
Flowmetry
Polaquiuria
Metastasis
present
absent
0,02
0,98
Ecography
Cultivation
PSA aux
CAT
IPSS
Hematuria
Dysuria
Rectal examination
stony
fibr_irr
normal
0,00
0,00
1,00
PSA total
Fever
Anaemia
Ganglion supra
Bone scanning
Loss of weight
Gleason
greater6
less6
zero
0,04
0,06
0,89
Pain
PSAI /PSAt
Hepatic afection
Constitutional syndrome


SPUTUM
no
yes
SPUTUM_COLOUR
purulent
non-purulent
PO2_FIO2
>_205
<=_205
LEUKOCYTOSIS
yes
no
RADIOLOGICAL_SIGNS
no
yes
MECHANICAL_VENTILATION
no
0-24_h
24-48_h
48-96_h
96-144_h
>_144_h
PNEUMONIA
no
yes
TEMPERATURE
>_38.5
normal
<_36.0


SPUTUM
no
yes
SPUTUM_COLOUR
purulent
non-purulent
PO2_FIO2
>=_205
<=_205
LEUKOCYTOSIS
yes
no
RADIOLOGICAL_SIGNS
no
yes
MECHANICAL_VENTILATION
no
0-24_h
24-48_h
48-96_h
96-144_h
>=_144_h
PNEUMONIA
no
yes
TEMPERATURE
>=_38.5
normal
<=_36.0
ANTIPYRETIC_DRUGS
no
yes

Before presenting any evidence, the probability of watchMovie is p_a.

The following findings are considered important (in order of importance):

- workDone results in a posterior probability of p_{b_1} for watchMovie.
- hasMoney results in a posterior probability of p_{b_2}
- hasFriend results in a posterior probability of p_b

Their influence flows along the following paths:

- workDone influences hasTime, which influences wat
- hasMoney influences watchMovie.
- hasFriend influences watchMovie.

Presenting the evidence results in a posterior probabil

The value **scirrheus** of node **Shape** is certain (P = 1.00).

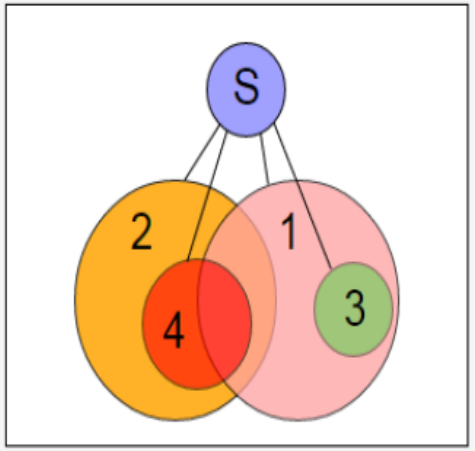
We were able to construct four arguments based on the evidence associated with the value **scirrheus** for node **Shape (S)**.
The arguments are ordered by how influential they are for the value of the node **Shape (S)**

- Argument 1: Node **Endosono-mediast** has value **no**
  Node **Bronchoscopy** has value **no**
  Node **Lapa-diagragm** has value **no**
  Node **CT-organs** has value **none**
  Node **X-fistula** has value **no**
  Node **CT-liver** has value **no**
  Node **X-lungs** has value **no**
  Node **CT-lungs** has value **no**
  Node **Endosono-wall** has value **T3**

- Argument 2: Node **Gastro-shape** has value **scirrheus**
  Node **Gastro-circumf** has value **circulair**
  Node **Gastro-length** has value **5 <= x < 10**
  Node **Weightloss** has value **x<10%**
  Node **Endosono-wall** has value **T3**
  Node **Endosono-truncus** has value **non-determ**
  Node **Endosono-loco** has value **yes**
  Node **Gastro-necrosis** has value **no**
  Node **X-fistula** has value **no**
  Node **Endosono-mediast** has value **no**
  Node **Gastro-location** has value **distal**

- Argument 3: Node **Gastro-shape** has value **scirrheus**

- Argument 4: Node **X-fistula** has value **no**
  Node **Gastro-necrosis** has value **no**



```
The following scenario(s) are
compatible with cold:
A. Cold and no cat hence no
   allergy                          0.47
   Other less probable
   scenario(s)                      0.06

The following scenario(s) are
incompatible with cold:
B. No Cold and cat causing
   allergy                          0.48

Scenario A is about as likely as
scenario B (0.47/0.48)
because cold in A is a great deal
less likely than no cold in B
(0.08/0.92),
although no cat in A is a great deal
more likely than cat in B (0.9/0.1).

Therefore cold is slightly more
likely than not (p=0.52).
```

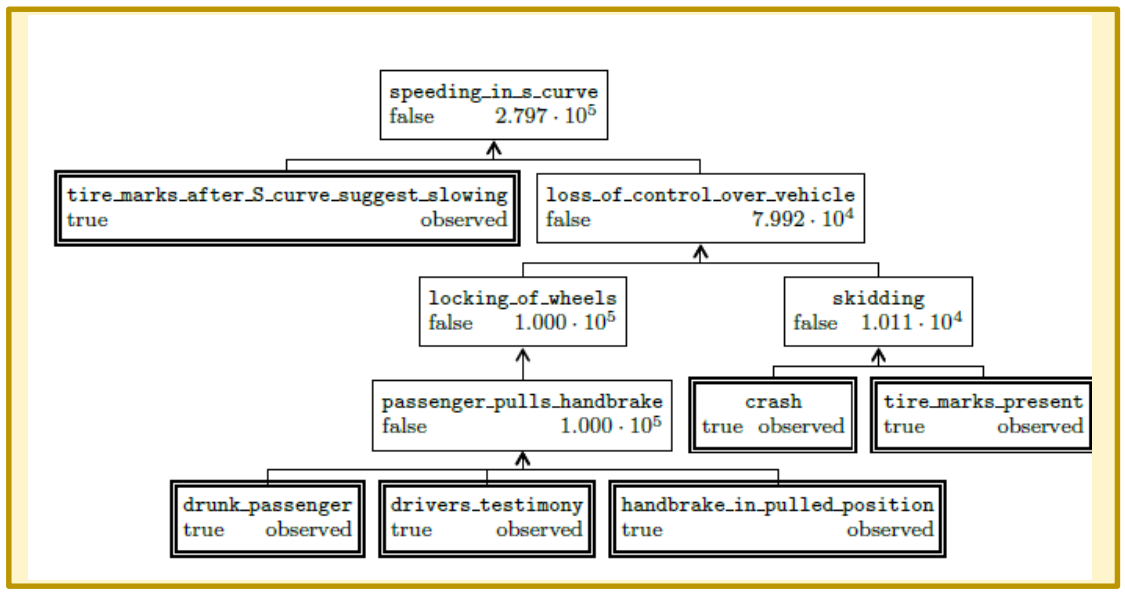
Scenario 2: Sylvia and Tom committed the burglary. (prior probability: 0.0001, posterior probability: 0.2326)

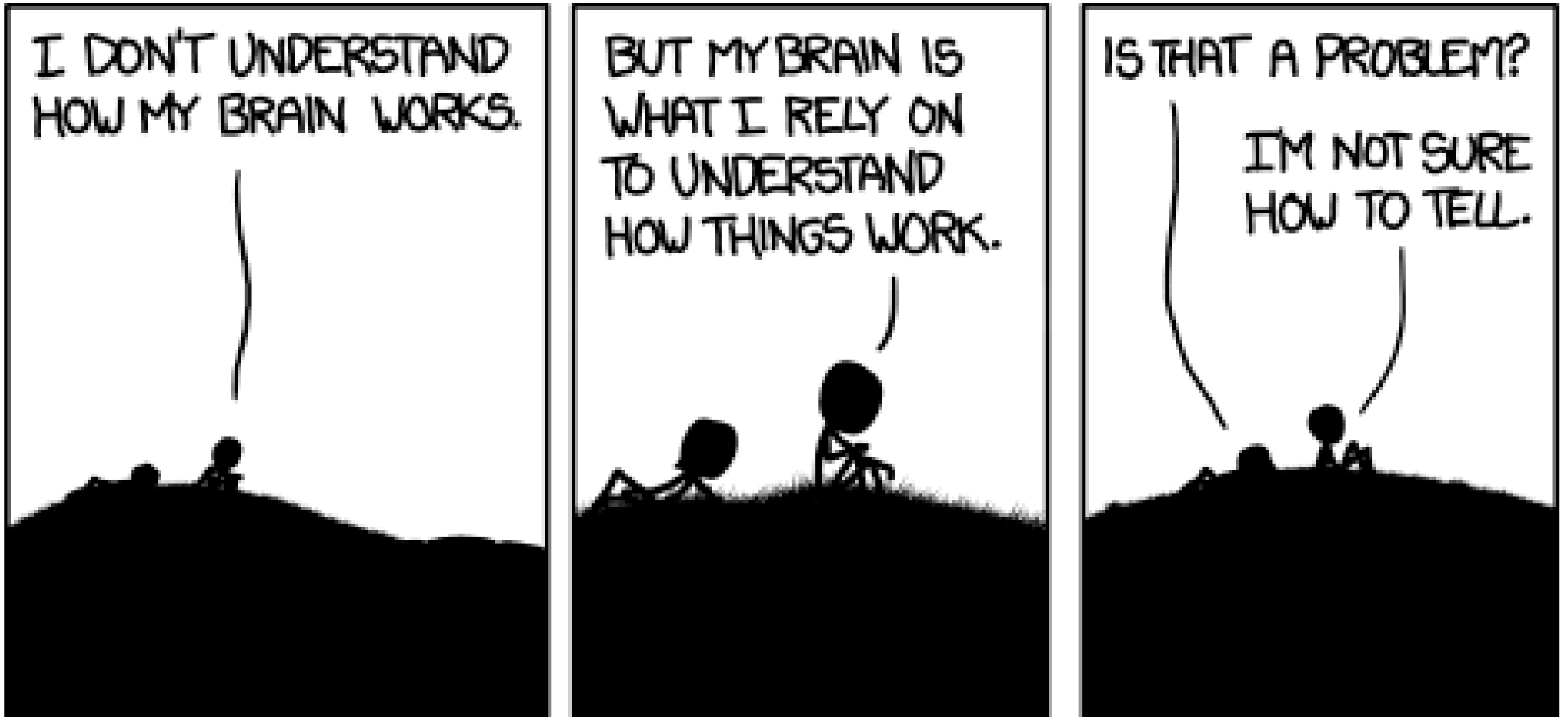
> **Scenario: Sylvia and Tom committed the burglary**: Sylvia and Tom had debts and a window was already broken. Then, Sylvia and Tom climbed through the window. Then, Tom stole a laptop.

Scenario 2 is complete and consistent. It contains the evidential gap 'Sylvia and Tom had debts' and the supported implausible element 'A window was already broken'.

Evidence for and against scenario 2:

* Broken window: moderate evidence to support scenario 2.
* Statement: Tom sold laptop: moderate evidence to support scenario 2.
* Testimony: window was already broken: weak evidence to support scenario 2.
* All evidence combined: very strong evidence to support scenario 2.




I DON'T UNDERSTAND HOW MY BRAIN WORKS.
BUT MY BRAIN IS WHAT I RELY ON TO UNDERSTAND HOW THINGS WORK.
IS THAT A PROBLEM?
I'M NOT SURE HOW TO TELL.

"Artificial intelligence is nowhere near matching the cognitive ability of an infant [...] unable to compete with the average human four-year-old."

John Brockman: POSSIBLE MINDS: 25 Ways of Looking at AI.
Penguin Press, 2019 John Brockman.

Four-Year-Old Boy Drives Great Grandfather's Car A Mile To Buy Sweets

DOMINIC SMITHERS in NEWS

Last updated 21:23, Saturday 15 June 2019 BST


DIPLOMA

Sources of images:

- Explainable AI: : https://blog.global.fujitsu.com/fgb/2019-08-01/why-ai-got-the-answer-explainable-ai-showing-bases/
- Asia networks: Daly, Shen, Aitken: Learning Bayesian networks: Approaches and issues. The Knowledge Engineering Review 26 (2011),
  and https://www.bayesserver.com/docs/introduction/bayesian-networks
- Elvira: Lacave, Luque, Díez: Explanation of Bayesian Networks and Influence Diagrams in Elvira. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 2007
- VaP network: PhD thesis Stefan Visscher
- Brain Cartoon: https://xkcd.com/1163/
- Autonomous car: https://theconversation.com/to-get-the-most-out-of-self-driving-cars-tap-the-brakes-on-their-rollout-88444
- Graduated robot: depositphotos.com image ID 200717136

Utrecht University

DISCLAIMER

The information in this presentation has been compiled with the utmost care,
but no rights can be derived from its contents.