TEACHING AI THE ART OF MAKING MORAL DECISIONS

PROF. TAD GONSALVES, PHD

DEPT. OF INFORMATION & COMMUNICATION SCIENCES

FACULTY OF SCIENCE & TECHNOLOGY

SOPHIA UNIVERSITY

SOPHIA UNIVERSITY

TOKYO

GONSALVES AI LAB (Yotsuya 2019)

www.gonken.tokyo
/

Deep Learning Applications Research Topics

SELF DRIVING CARS & DRONES

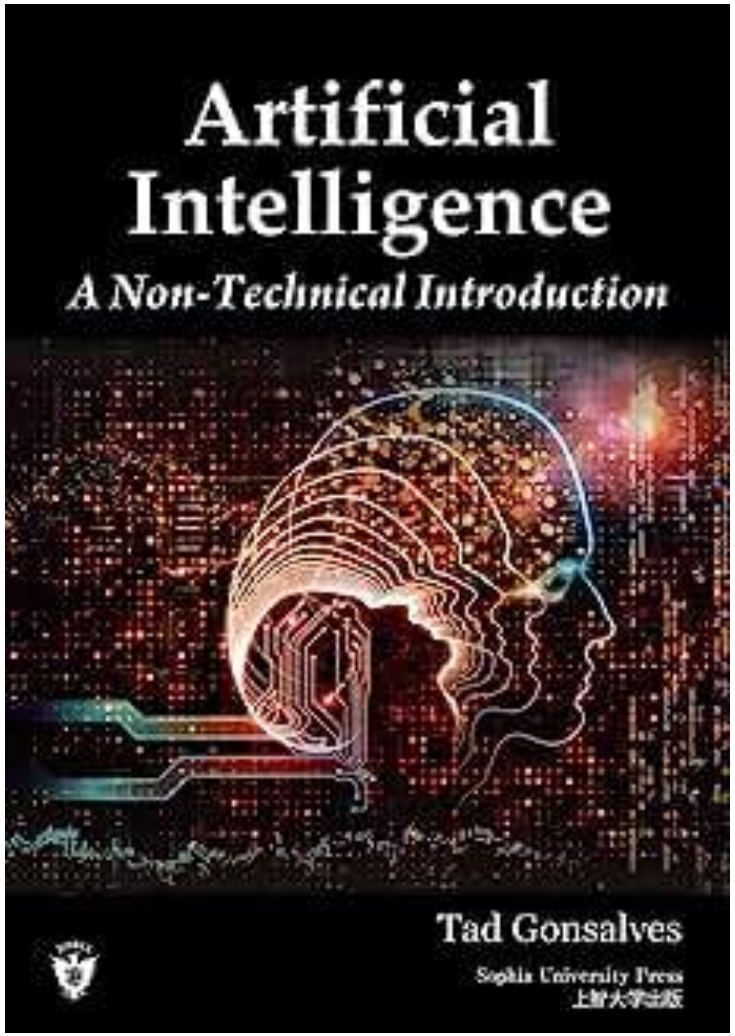
PC GAMES & PUZZLES

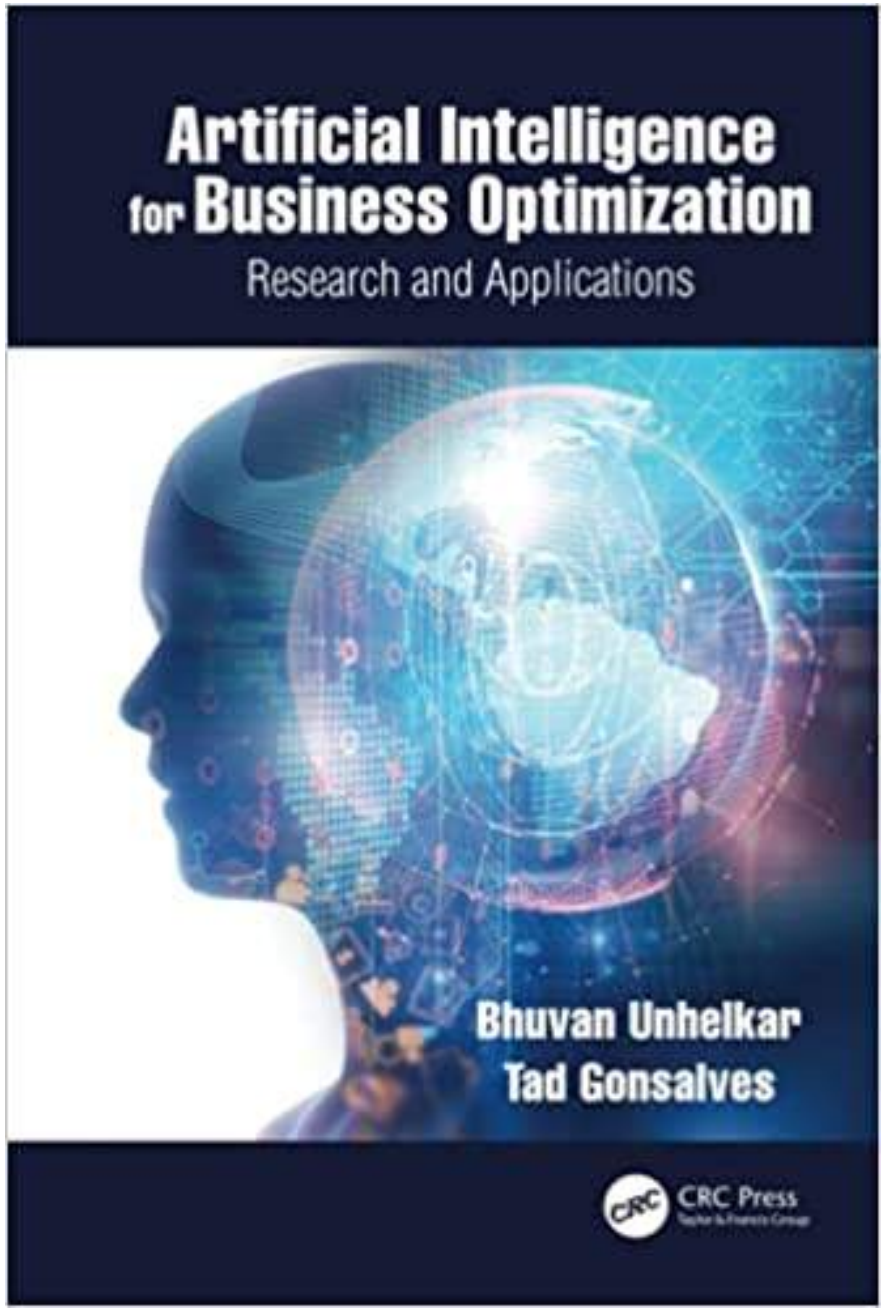
NATURAL LANGAUGE PROCESING (SPANISH)

DIGITAL ART & MUSIC

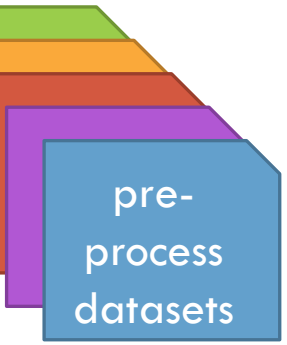
FINANCE

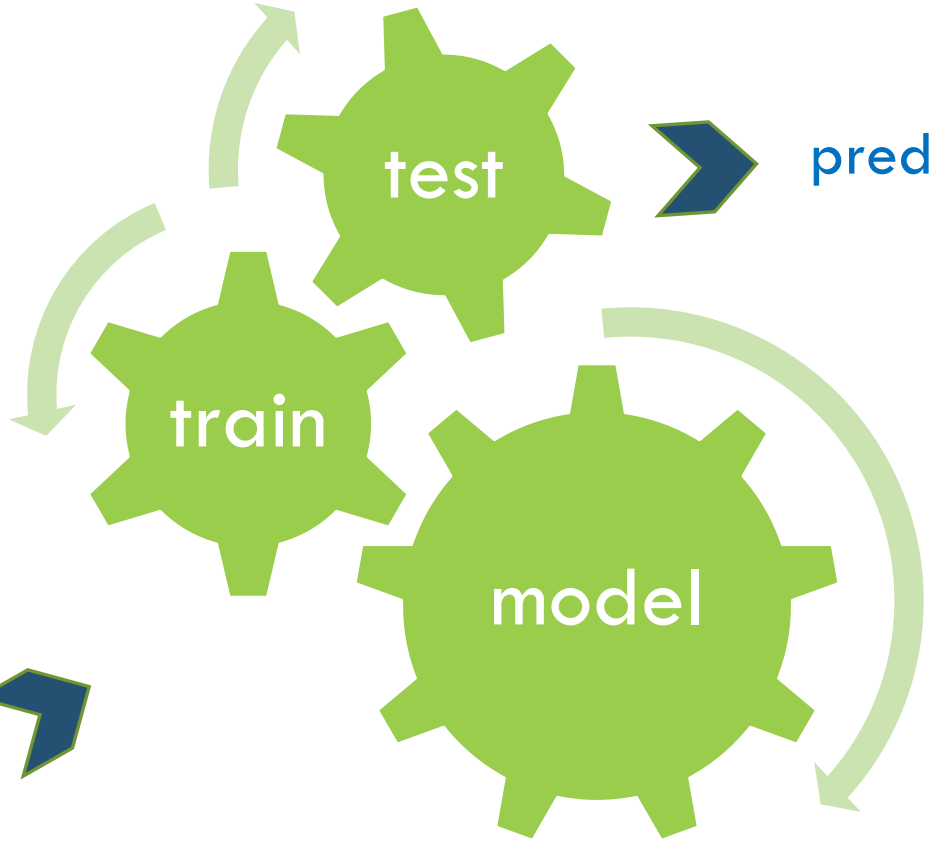
TWO FAMOUS BOOKS





AI MACHINE LEARNING PIPELINE





predictions

biased
tainted
not transparent
not explainable
not accountable
……

AI in its current form does not have the least idea of ethics!!

EXPERT SYSTEM (1970 ~ 80)

CODIFY

FORMALIZE

use

KNOWELDGE (from domain experts)

RULE FORMAT: IF ( ) THEN ( ) BECAUSE ( )

AI PARADIGM
EXPERT SYSTEMS V/S MACHINE LEARNING

	USAGE	TRUST
EXPERT SYSTEMS	LOW	HIGH
MACHINE LEARNING	HIGH	LOW

THE PROBLEM

- Golden rule of ML: "GARBAGE IN, GARBAGE OUT!"
- Nobody has taught AI / ML algorithms what is right / wrong.
- All our research is about designing algorithms and tweaking them to produce accurate predictions given a dataset.
- How will the algorithms judge the predictions are ethically right?

Rules and
regulations
(external)

COMPUTING PARADIGM

Catholic univ contribution

Stage 3: AEI

Stage 2: AI

Stage 1: noise, data, information

wisdom

meta-knowledge

knowledge

information

data

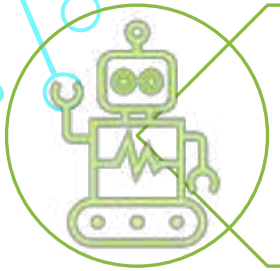
noise

HUMAN ETHICS

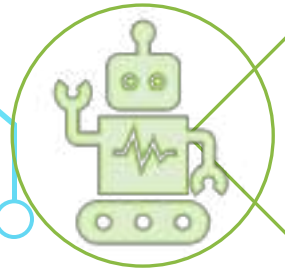
- 1) **Knowledge** (acquired over the years)
  - (a) Ethical principles of profession
  - (b) Domain knowledge of the problem
  - (c) Consequences
- 2) **Experience** (learnt through trial and error)
  - (a) Fuzzy nature of moral situations
  - (b) Uncertainty & Contradictions
  - (c) Case based reasoning
- 3) **Empathy** (conscience)
  - (a) Developed through education and upbringing
  - (b) Feeling rather thinking when making moral decisions
  - (c) Conscience acts as the final arbitrar

Outstanding contribution of Catholic education

AI CONSCIENCE



Keep revising and updating knowledge



Acquire meta-knowledge through fuzzy Machine Learning

Artificial Emotional Intelligence (AEI)

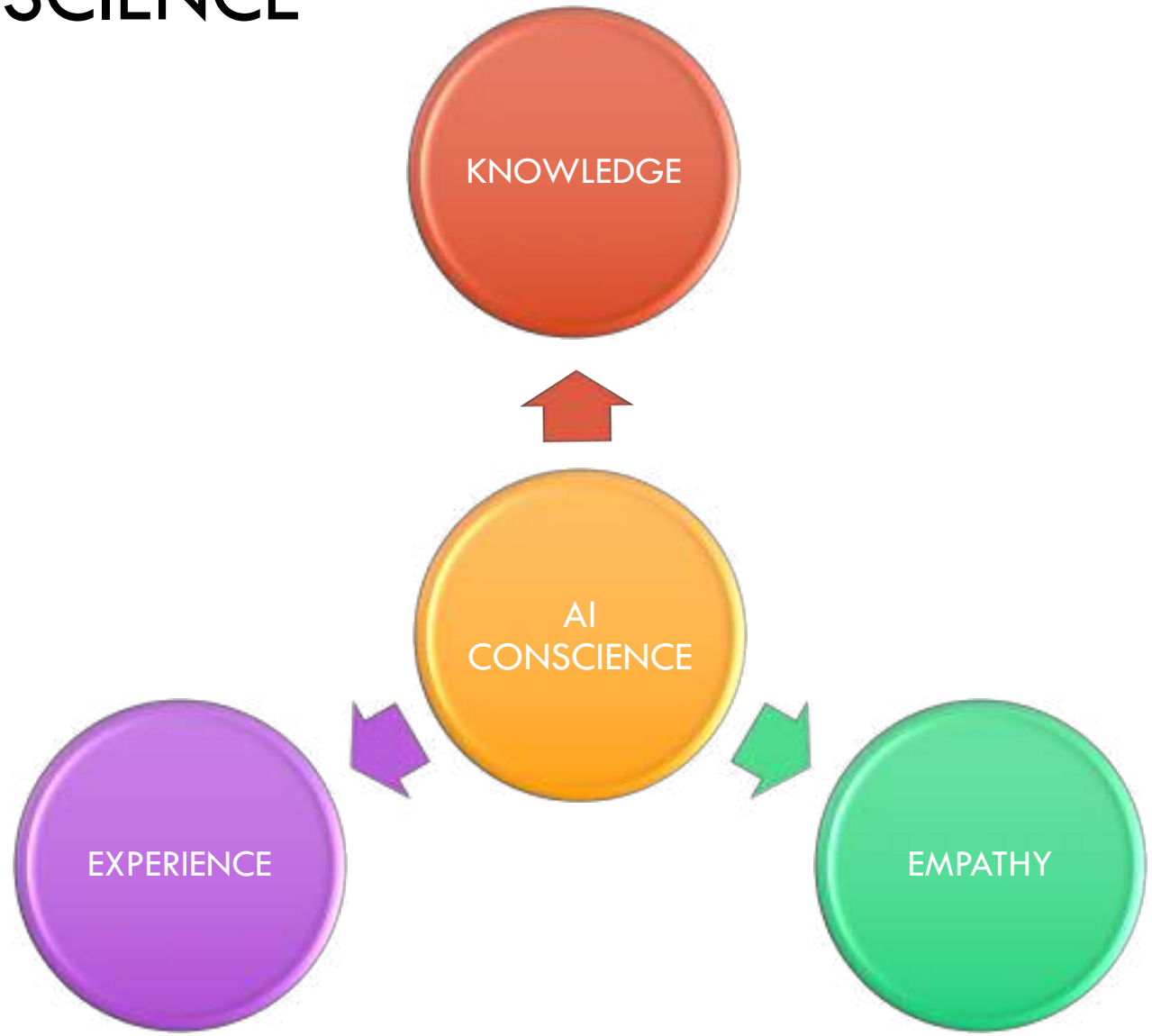
HUMAN CONSCIENCE

knowledge

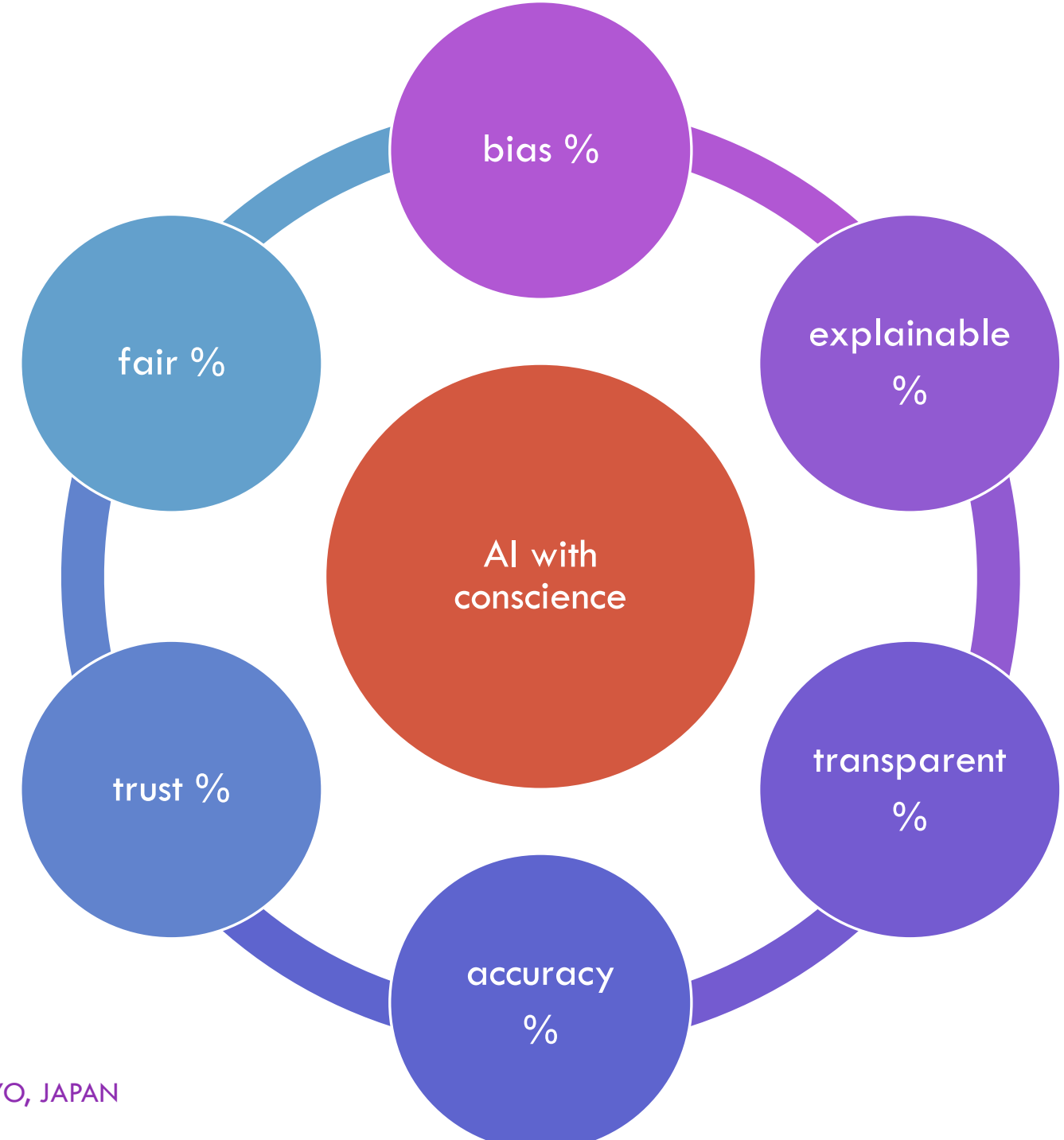
experience

empathy

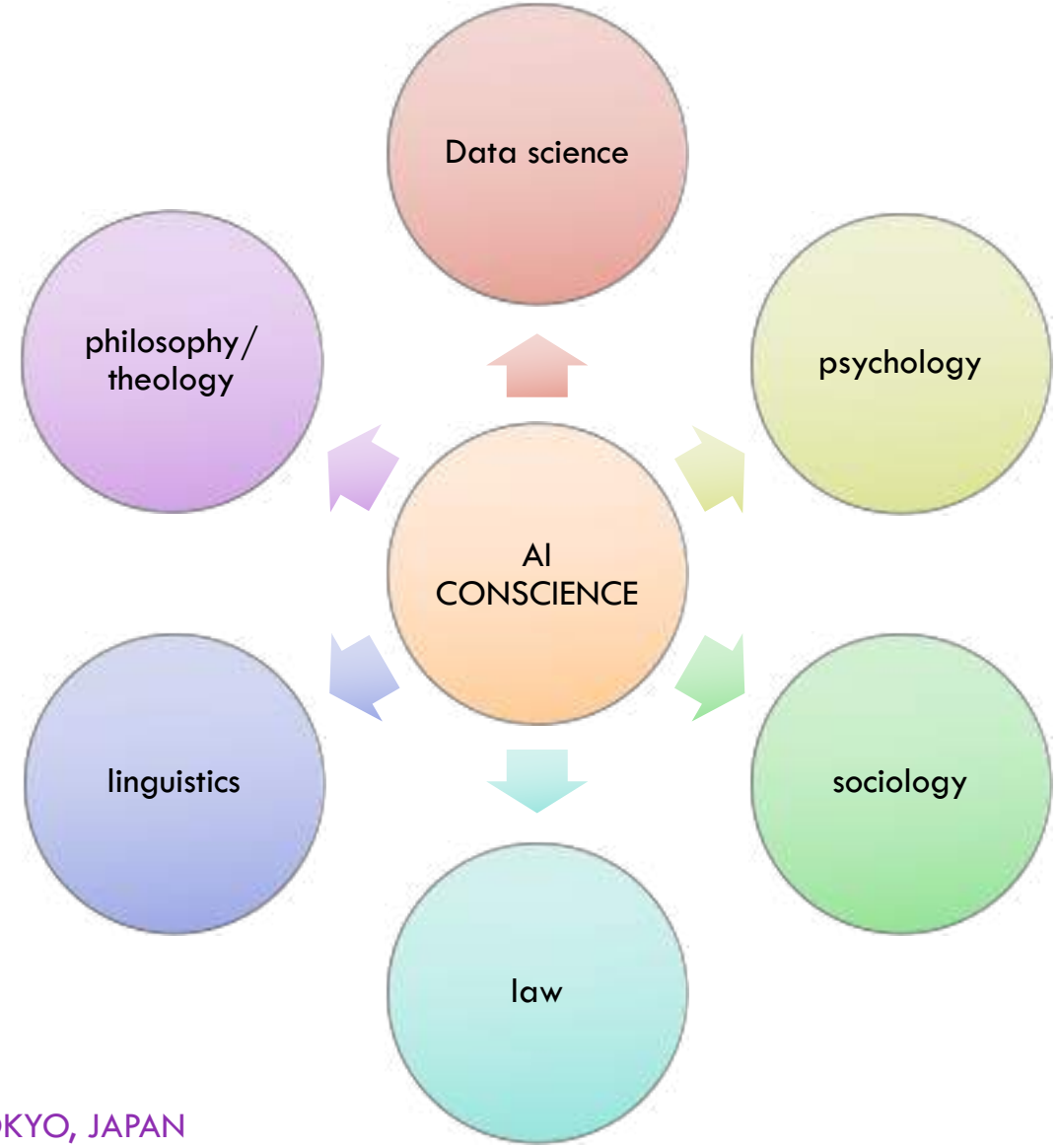
AI CONSCIENCE



Rules and
regulations
(external)


bias %
explainable %
transparent %
accuracy %
trust %
fair %
AI with conscience

MULTI-DISCIPLINARY AI CONSCIENCE



Thank you very much!

ご清聴ありがとうございます!

Grazie!