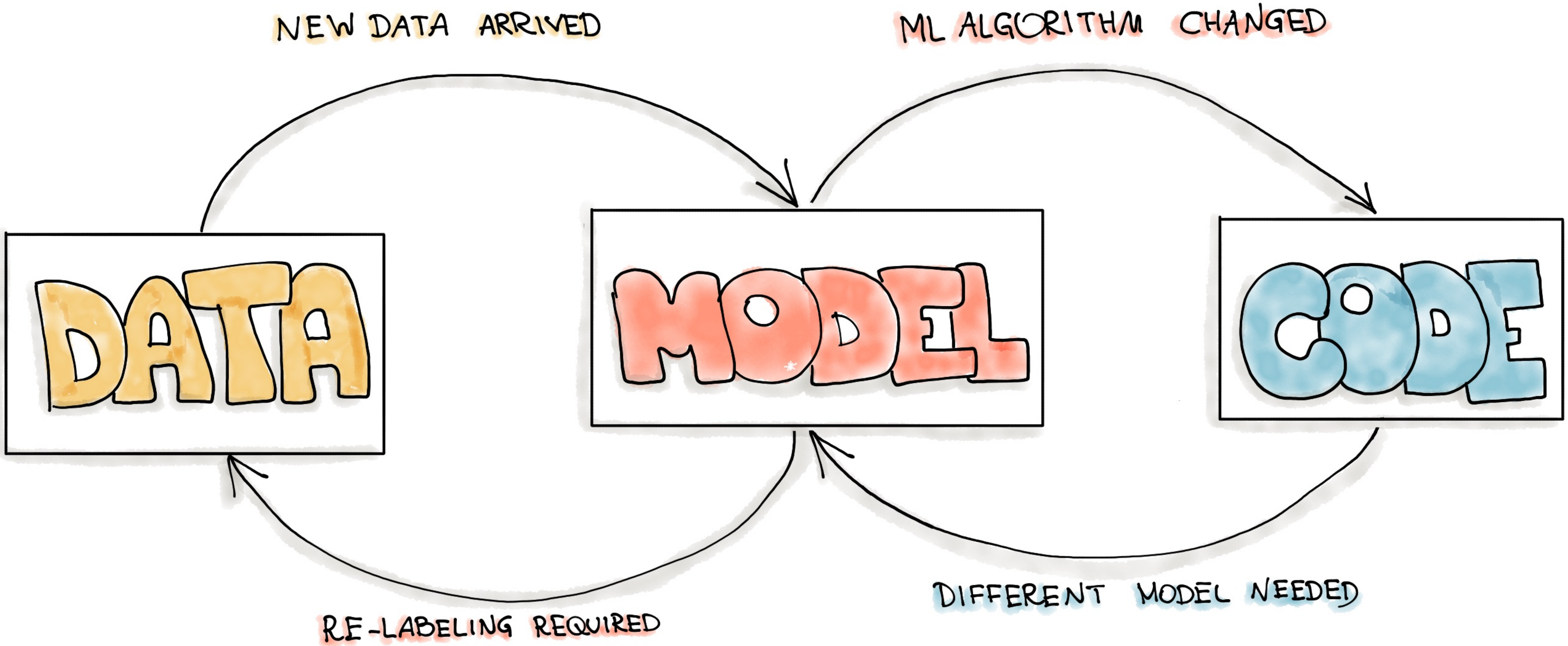
10 FOUNDATIONAL PRACTICES OF MACHINE LEARNING ENGINEERING

@visenger

PROGRAMMING vs. MACHINE LEARNING

3 LEVELS OF CHANGE



ML ENGINEERING HIERARCHY OF NEEDS

RESPONSIBLE AI

KNOW ML HIDDEN COSTS

DESIGN FOR ML

VERSIONING FOR ML

ML OPERATIONS

ML ALGORITMS SIMPLICITY

NO AI WITHOUT IA

ML PROJECT MANAGEMENT

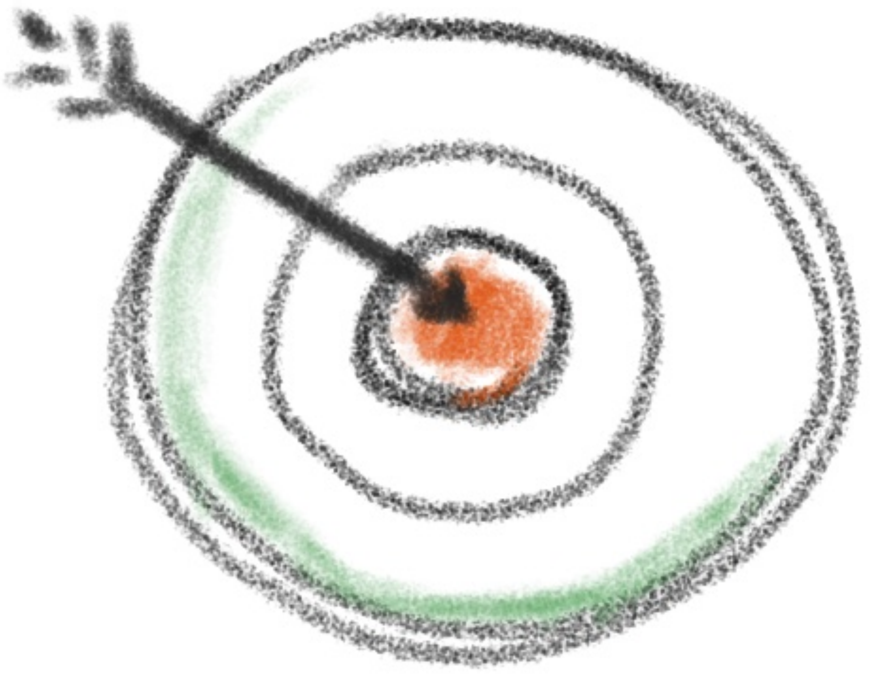
GET THE RIGHT TEAM / SKILLS

SOLVE ML PROBLEM

@visenger

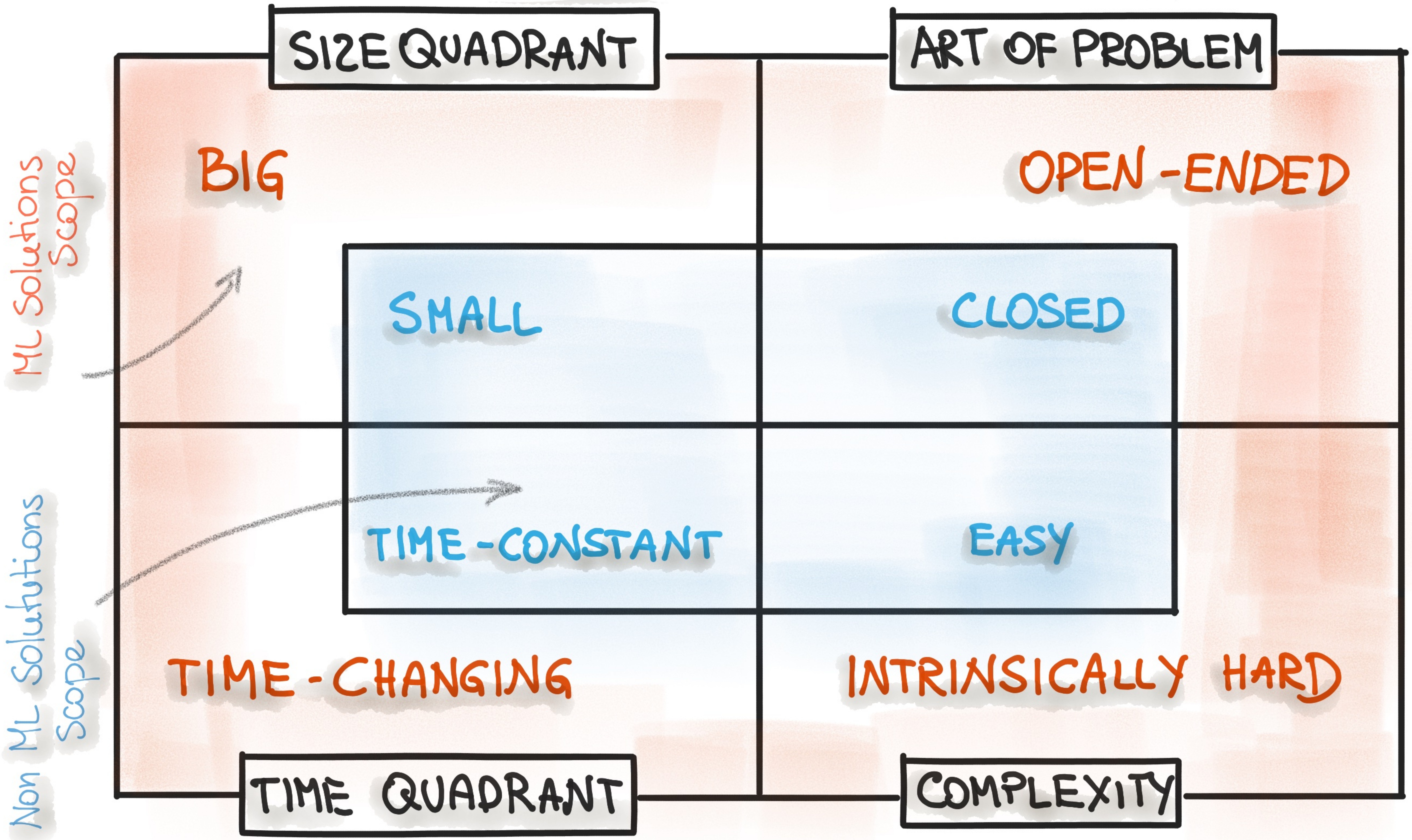
① AVOID ML-SOLUTIONISM

WHAT IS THE RIGHT PROBLEM FOR AN ML-SOLUTION ?



@visenger

① AVOID ML-SOLUTIONISM

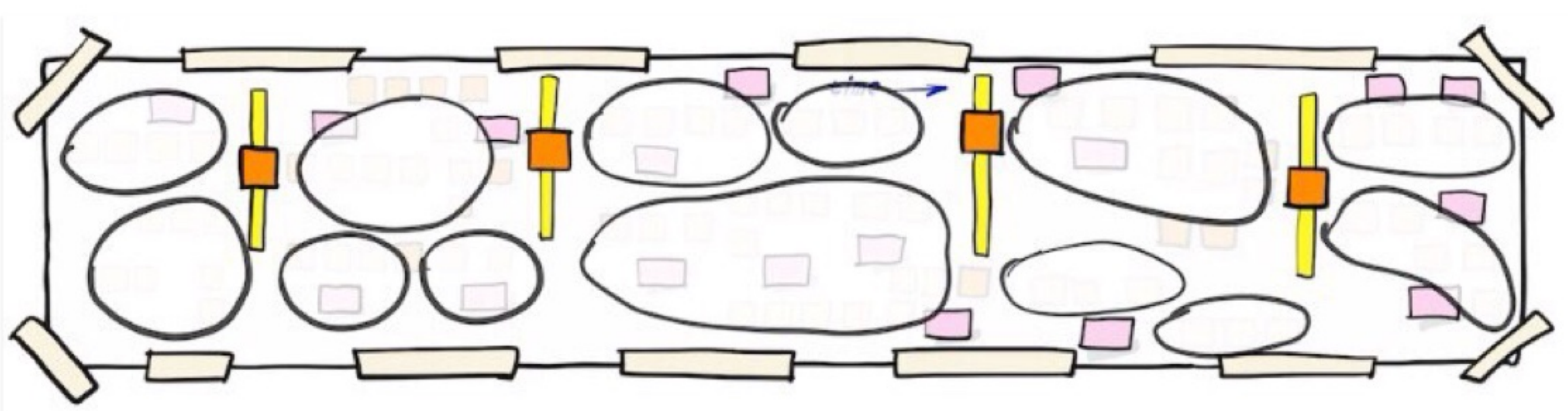


@visenger

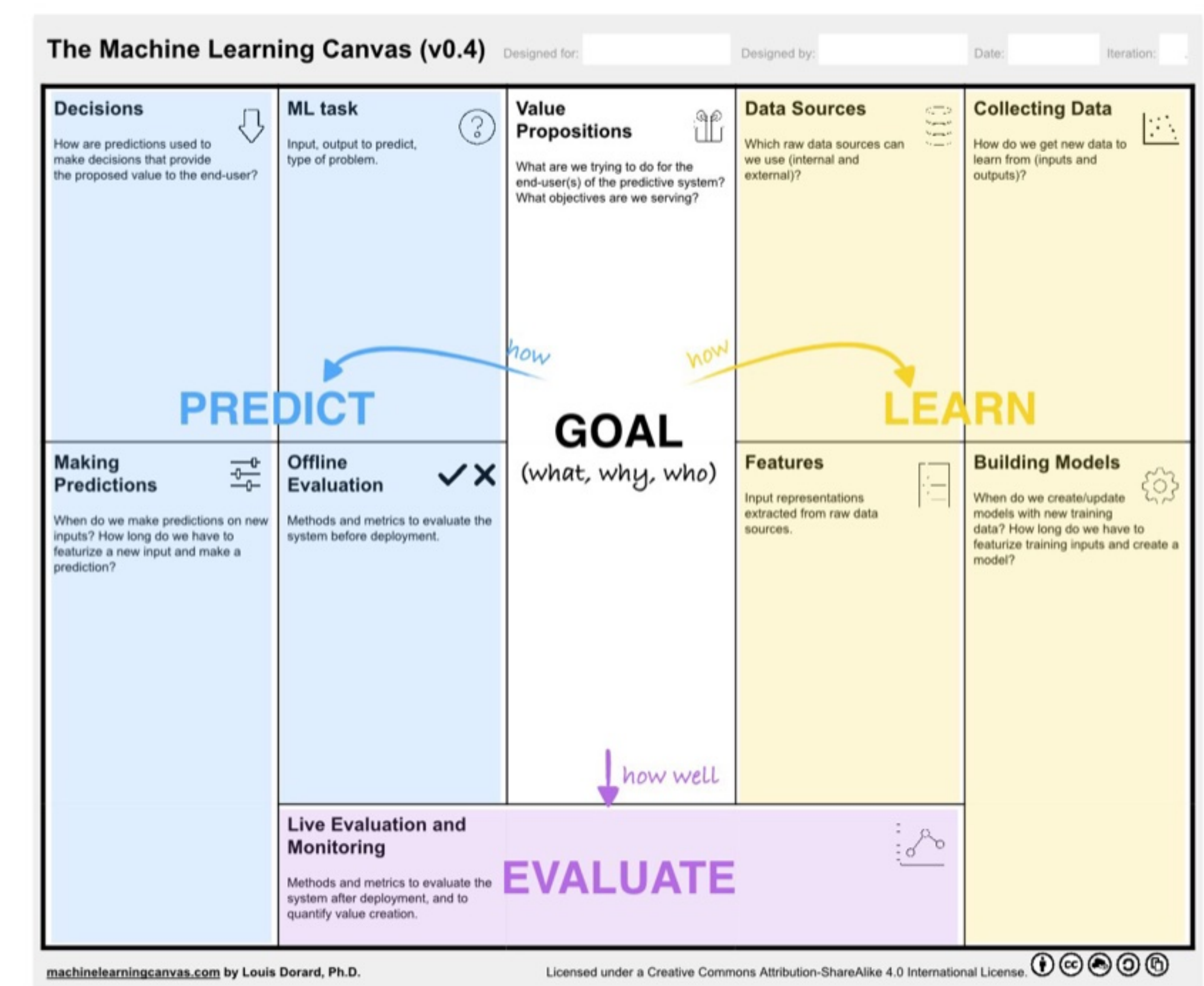
① AVOID ML-SOLUTIONISM

FIND THE OPPORTUNITY

STRUCTURE PROJECT



(picture by A. Brandolini)



DDD KNOWLEDGE CRUNCHING

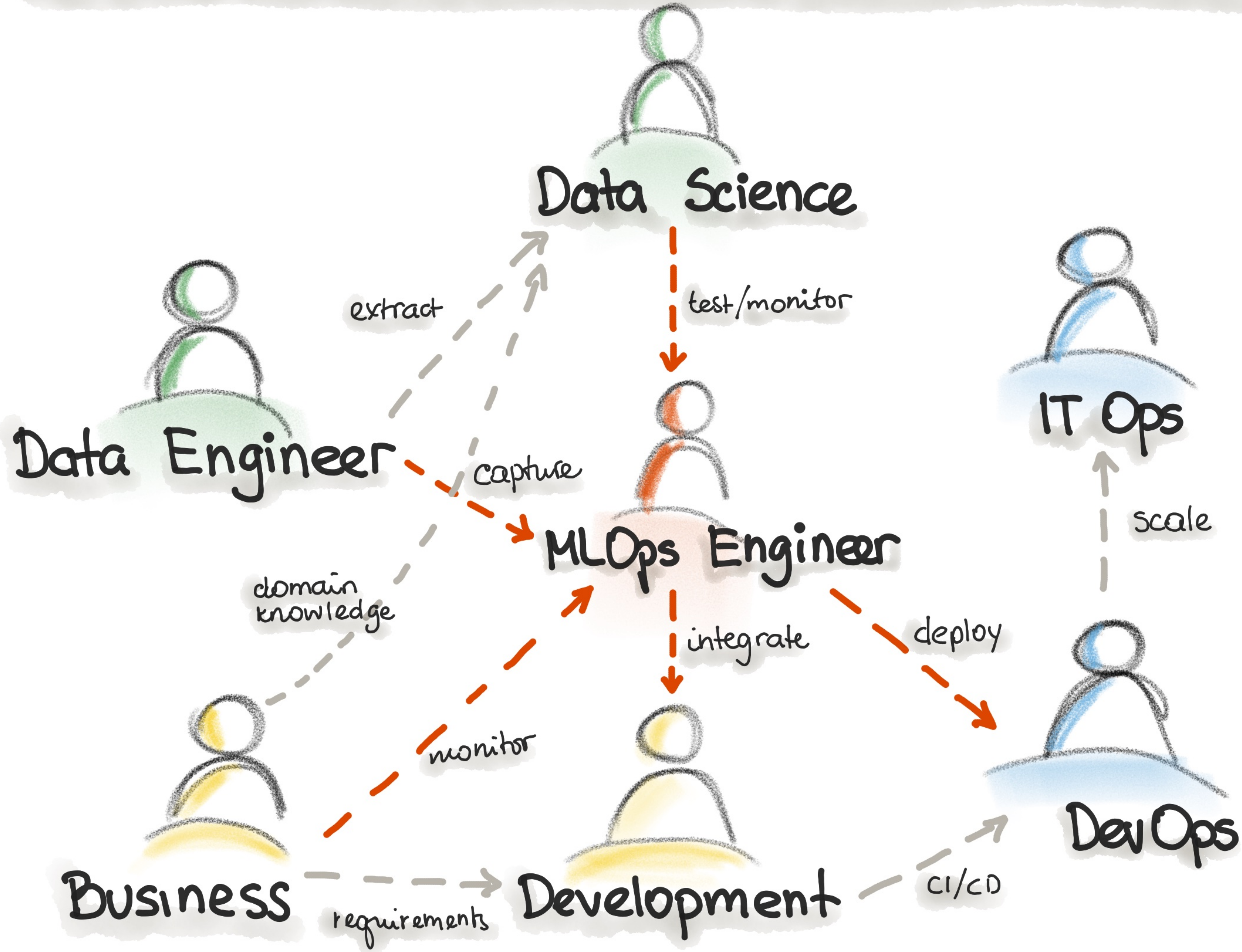
METHODS

EVENT STORMING

MACHINE LEARNING DESIGN CANVAS

@visenger

② GET THE RIGHT ML-TEAM



from "MLOps: Operationalizing Data Science" by D. Sweenor et al.

@visenger

3 ML-PROJECT ≠ SE-PROJECT

AIM	FINDING INSIGTHS IN THE DATA	IMPLEMENTING A SOLUTION FOR A SPECIFIC REQUIREMENT
FUNCTIONALITY	A MODEL IS NEVER 100% ACCURATE AND ALWAYS CHANGES POST-DEPLOYMENT PLANNING	SOFTWARE EITHER WORKS OR IT DOESN'T
TRACKING PROGRESS	TRACKING DELIVERY: - ML MODEL - NEW LIBRARY - NEW ALGORITHM	MEASURING PRODUCT DELIVERY

@visenger

③ ML-PROJECT ≠ SE-PROJECT

Problems

AD-HOC

NOT REPEATABLE

NOT SUSTAINABLE

NOT ORGANIZED

NOT DOCUMENTED

@visenger

③ ML-PROJECT MANAGEMENT

CRISP-ML(Q)

2020

CROSS-INDUSTRY STANDARD PROCESS FOR ML APPLICATIONS WITH QUALITY ASSURANCE

ML Project Phases

1. Business & Data Understanding
2. Data Preparation
3. Modeling
4. Evaluation
5. Deployment
6. Monitoring & Maintenance

TDSP

TEAM DATA SCIENCE PROCESS

4 COMPONENTS

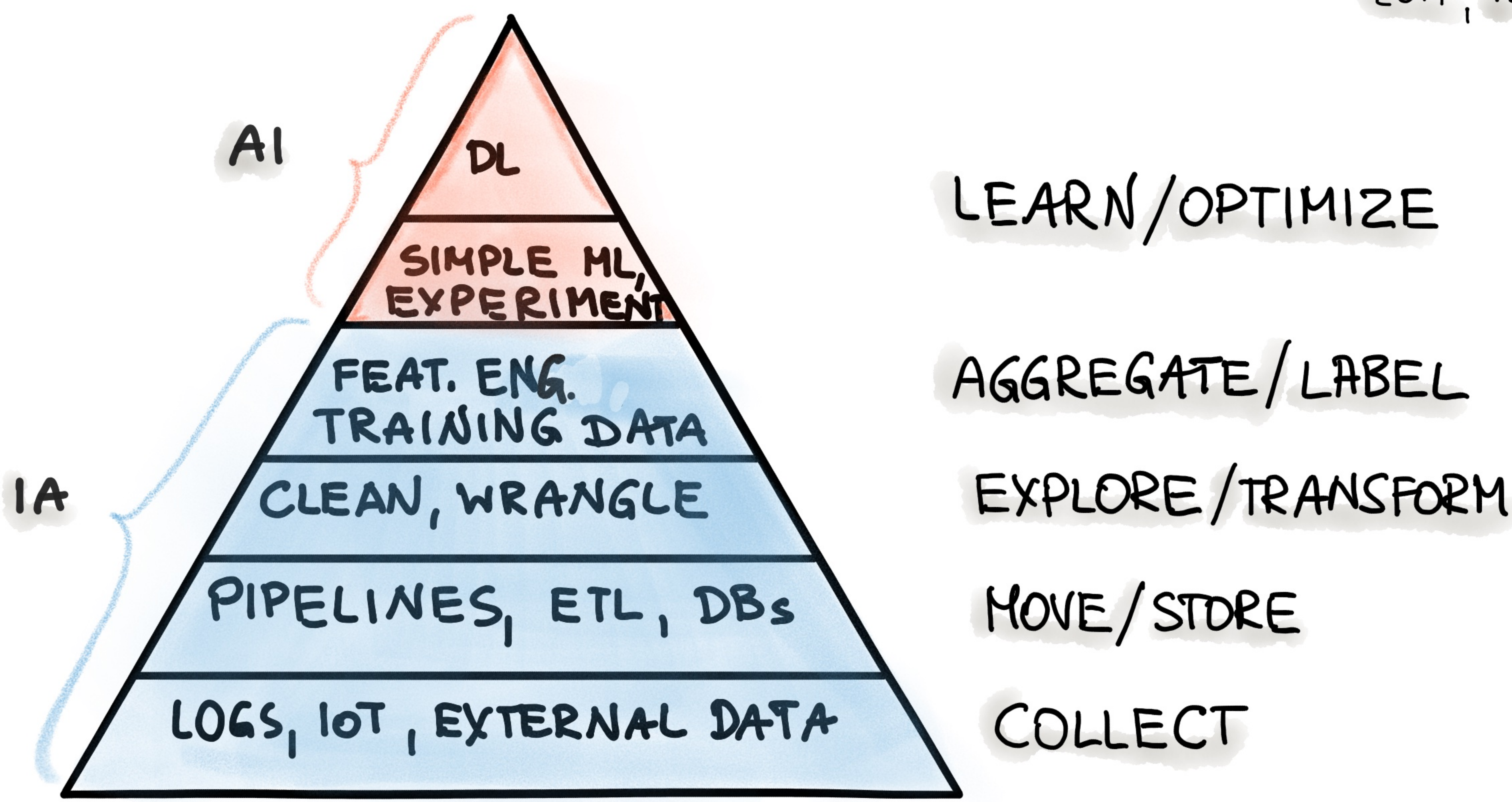
1. A Data Science Lifecycle
2. A Standardized Project Structure
3. Infrastructure & Resources
4. Tools & Utilities

@visenger

④ "THERE IS NO AI WITHOUT IA"

INFORMATION ARCHITECTURE

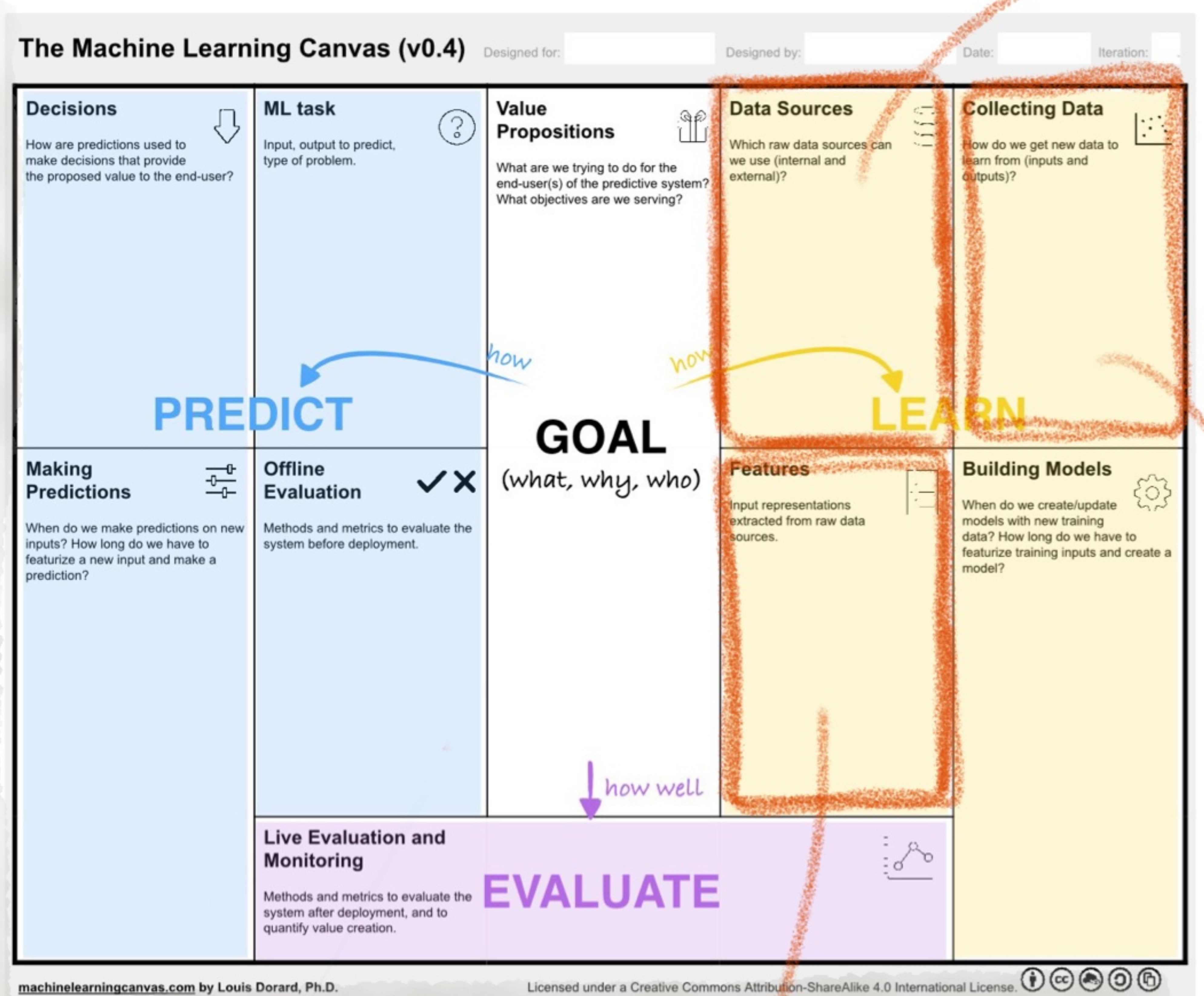
Seth Earley,
2017, IEEE Software



Monica Rogati, 2017

@visenger

④ "THERE IS NO AI WITHOUT IA"



WHICH RAW DATA SOURCES CAN WE USE TO TRAIN OUR ML MODEL?

HOW DO WE GET NEW DATA TO RE-TRAIN OUR MODEL?

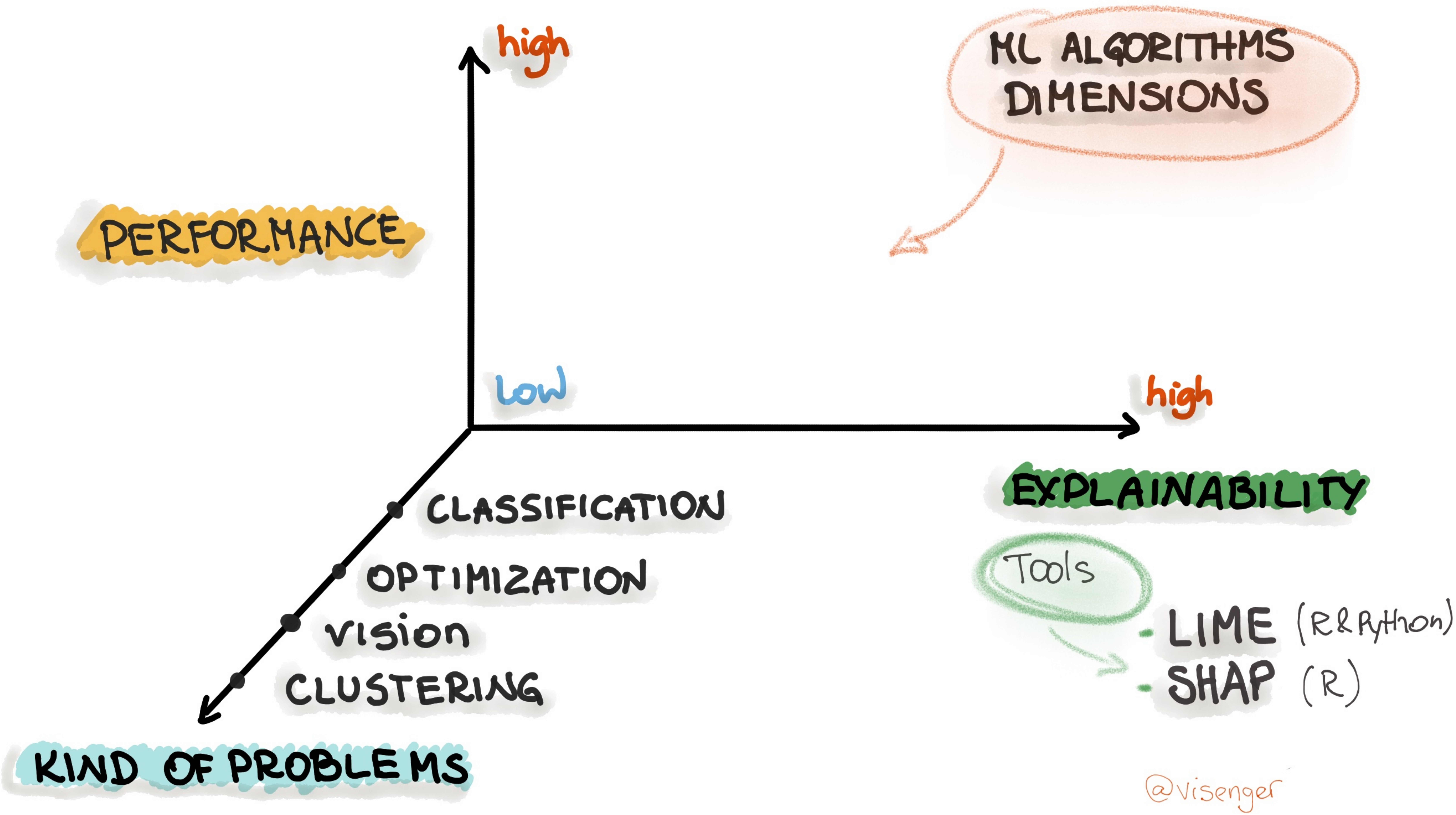
WHAT FEATURES DO WE NEED?

(Features Store)

@visenger

④ "THERE IS NO AI WITHOUT IA"

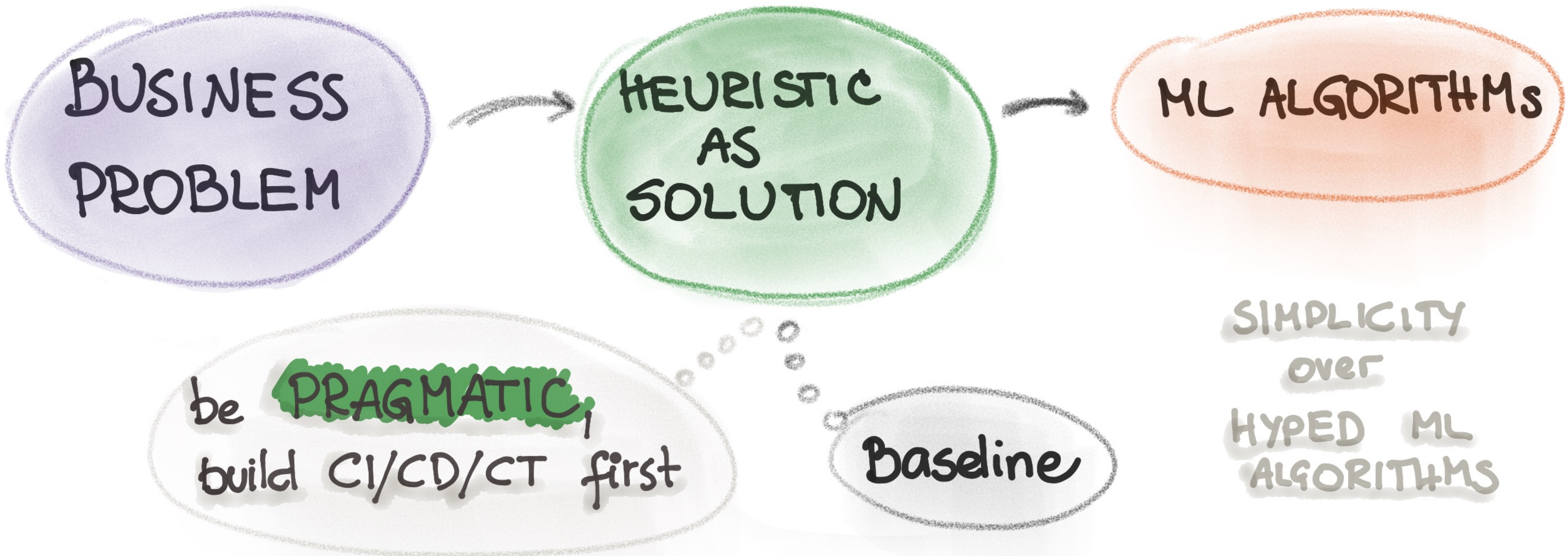
5 COMMON SENSE ML over EN VOGUE ML



5 COMMON SENSE ML over EN VOGUE ML

5 COMMON SENSE ML over EN VOGUE ML

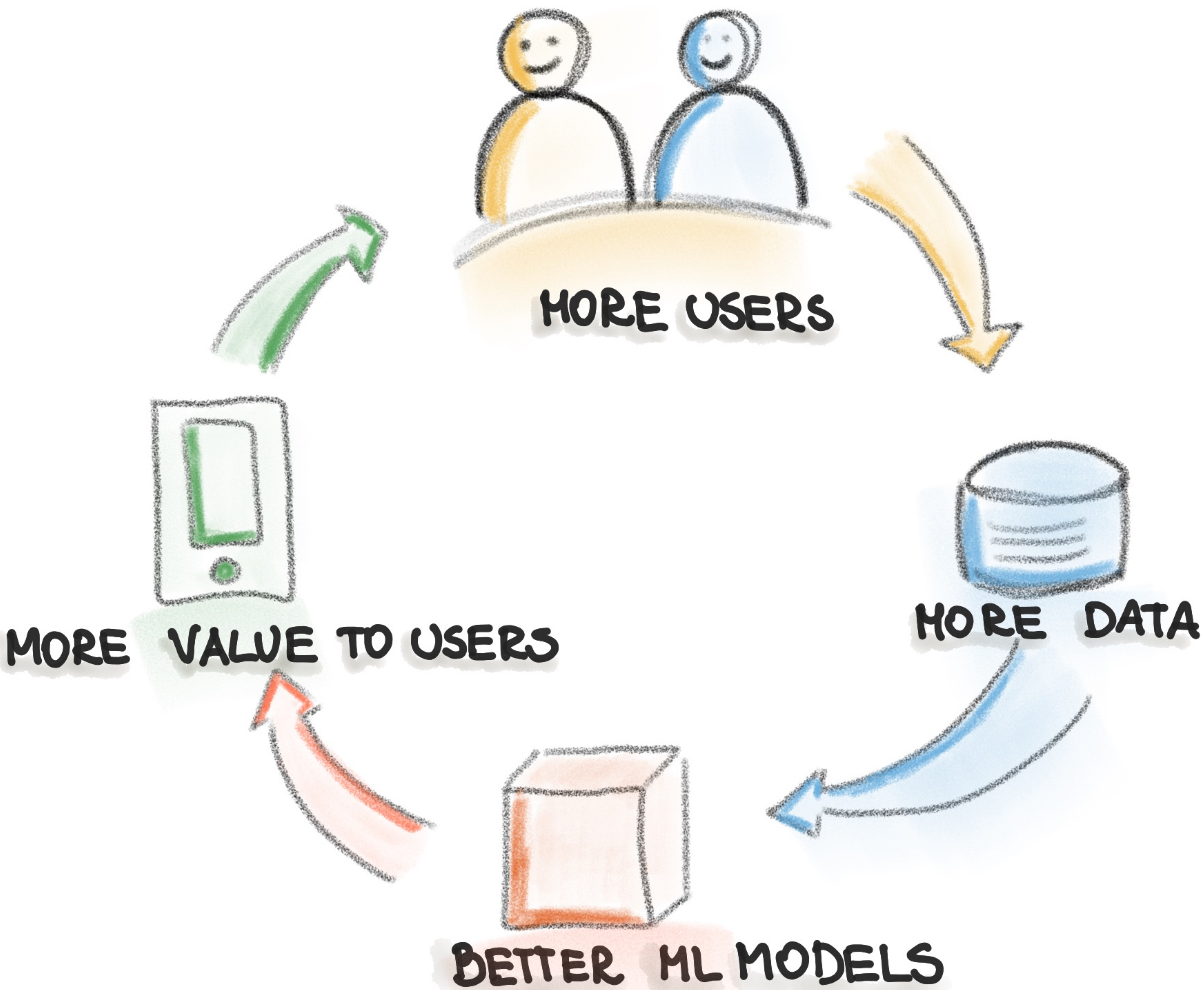
NO-ML-SOLUTION FIRST



M. Zinkevich "RULES OF MACHINE LEARNING"

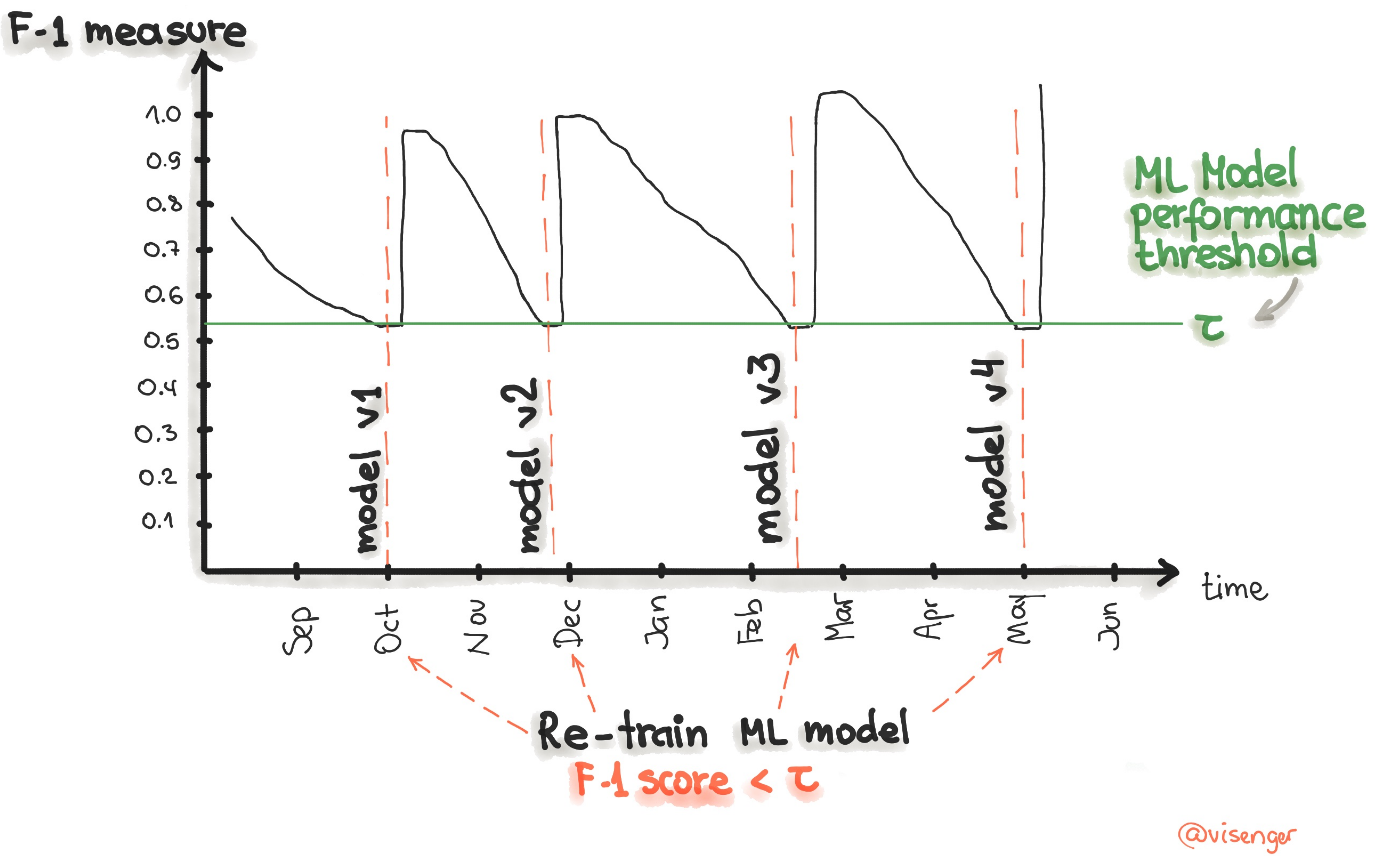
@visenger

⑥ THE FLYWHEEL OF ML SYSTEMS

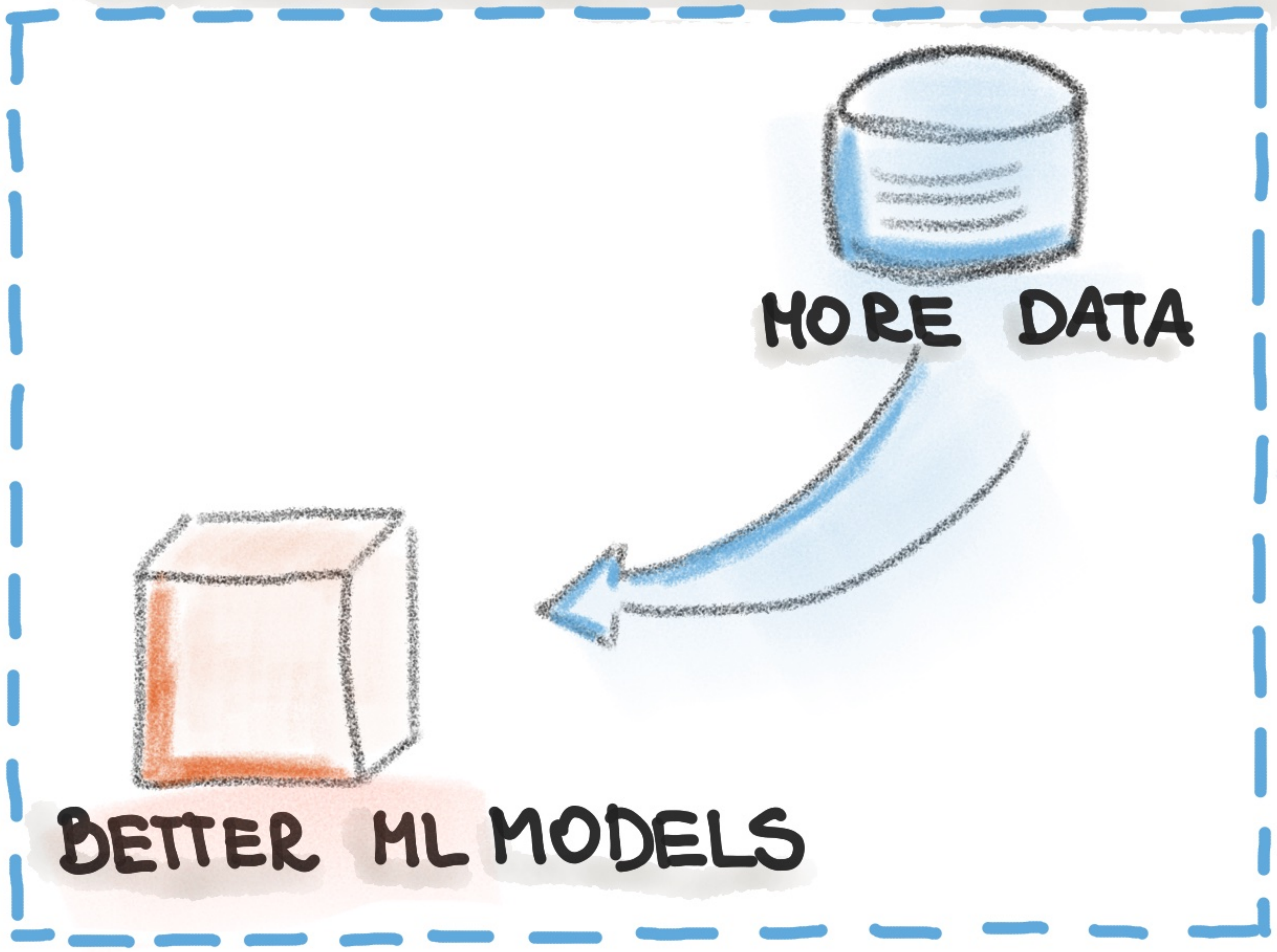


@visenger

ML MODEL DECAY MONITORING



⑥ THE FLYWHEEL OF ML SYSTEMS



be PRAGMATIC,
build CI/CD/CT first

IMPROVING ML MODELS

RE-TRAIN

EXCHANGE

REPLACE
WIHT
HEURISTICS

@visenger

⑥ THE FLYWHEEL OF ML SYSTEMS

ML TESTING TAXONOMY

WHERE

- DATA PIPELINE
- TRAINING PIPELINE

HOW

- OFFLINE (TRAINING)
- ONLINE (PREDICTION)

MONITORING

WHAT

NON-FUNCTIONAL (QUALITY) TESTING

- Efficiency
- Robustness
- Fairness
- Privacy
- Security
- Interpretability

FUNCTIONAL TESTING

- Correctness
- Model Relevance

@visenger

⑦ VERSIONING FOR ML SYSTEMS

CI/CT/CD TRIGGERS

CHECKPOINTS FOR

- RECOVERY,
- TRACEABILITY,
- DECISIONS JUSTIFICATION

@visenger

⑦ VERSIONING FOR ML SYSTEMS

CI/CT/CD TRIGGERS

ML MODELS

- TRAINING PIPELINE
- ML MODEL (OBJECT)
- HYPERPARAMETERS OPTIMIZATION
- EXPERIMENTS

DATA

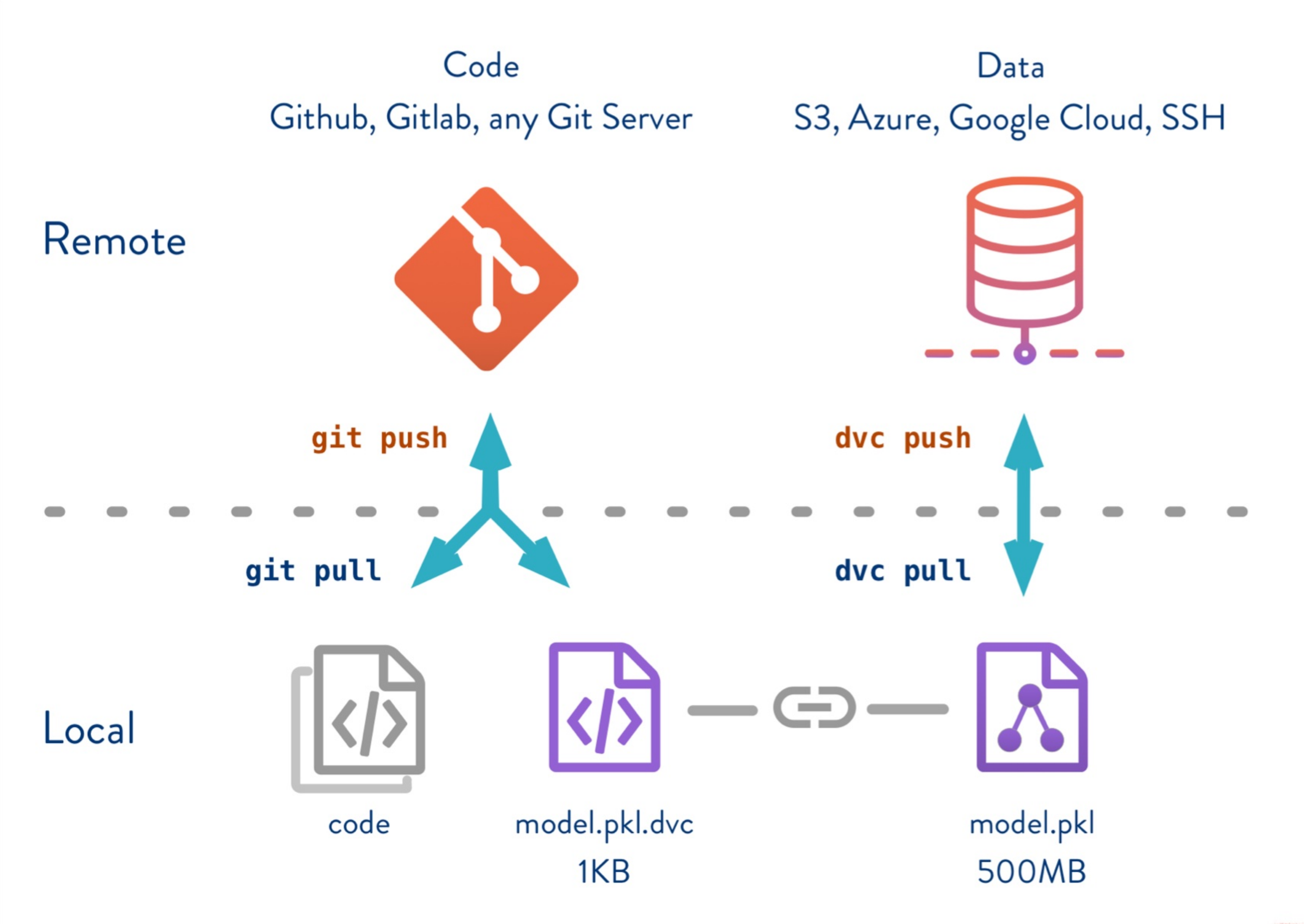
- DATA PREPARATION PIPELINE
- FEATURE STORES
- DATASETS
- METADATA

CODE

- APPLICATION CODE
- CONFIGURATION

@visenger

7 VERSIONING FOR ML SYSTEMS



TOOL

DVC.org

DATA VERSION CONTROL

@visenger

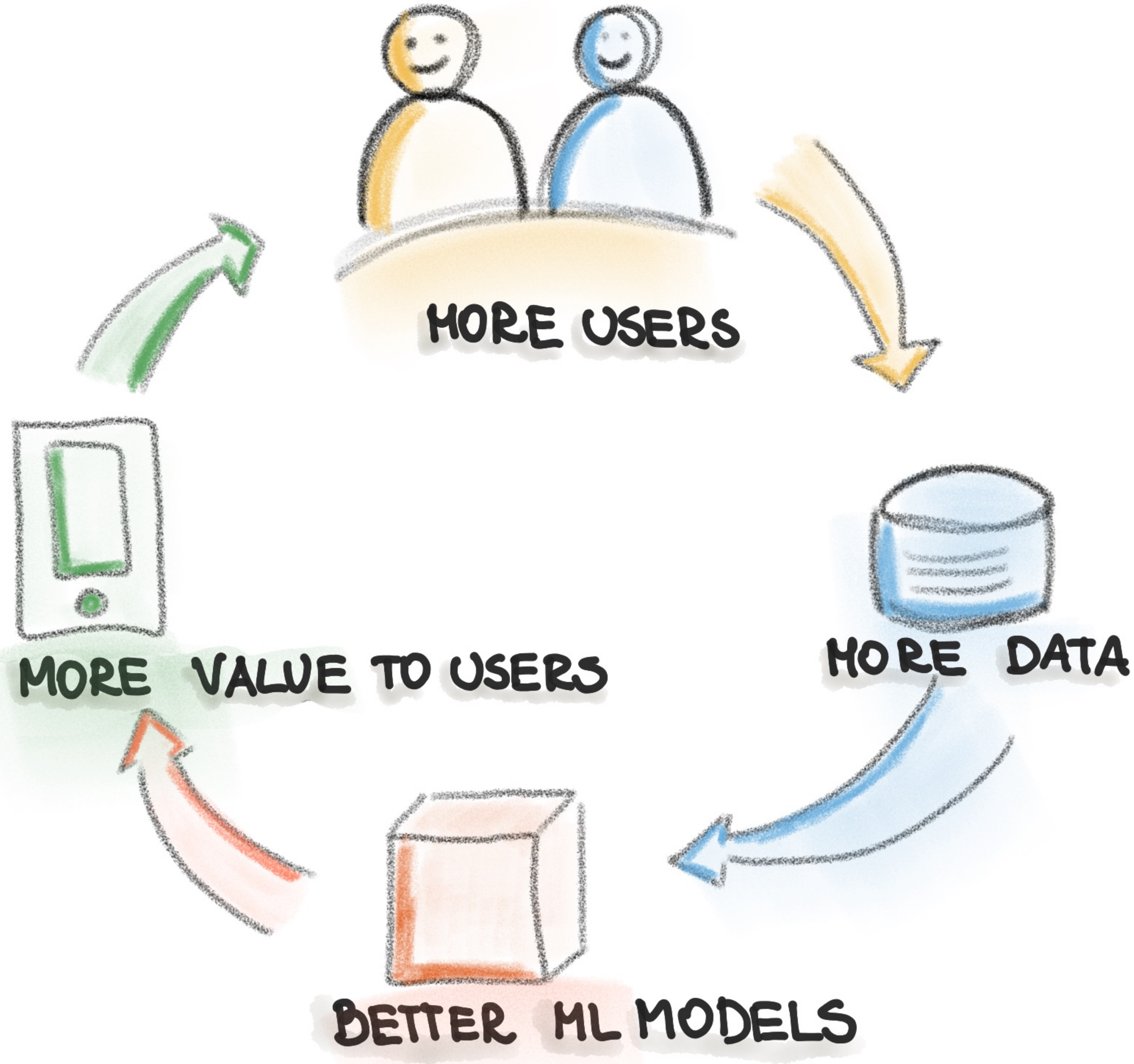
8 DESIGNING ML-SYSTEMS

ML MODELS SHOULD BE

- REPLACEABLE
- ADAPTABLE FOR CHANGE

⇓

LOOSE COUPLING

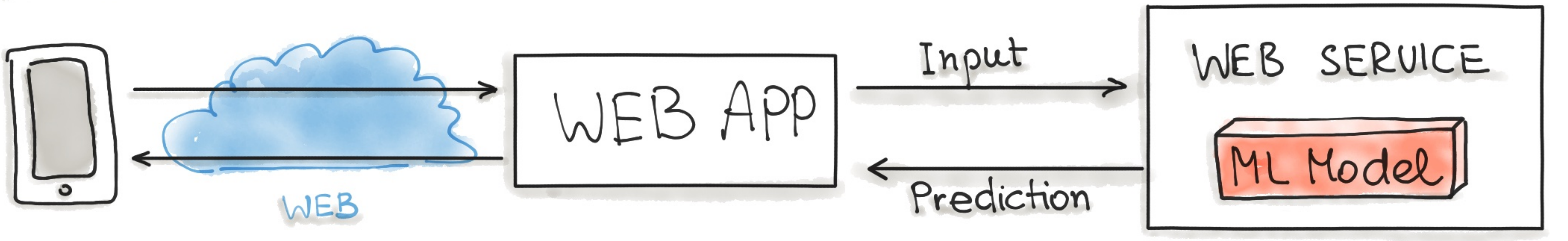


@visenger

⑧ DESIGNING ML-SYSTEMS

SERVING PATTERN FOR ML MODELS

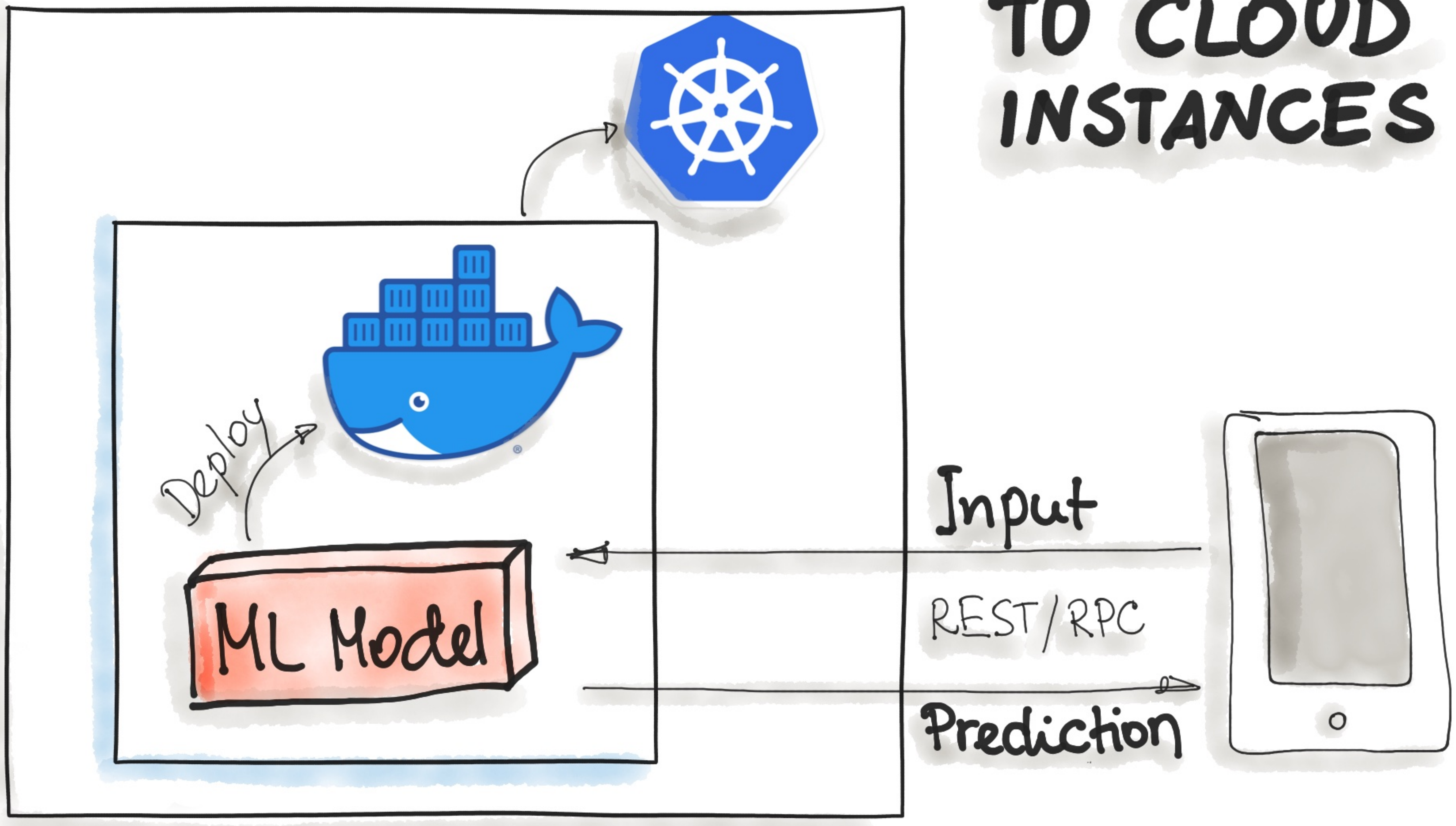
MODEL-as-SERVICE



@visenger

⑧ DESIGNING ML-SYSTEMS

INFRASTRUCTURE: ML MODEL DEPLOYMENT TO CLOUD INSTANCES



@visenger


MLOps
MODEL DEVELOPMENT
ML OPERATIONS
Model analysis
Data analysis
Experiments
Source repository
CI/CD Stage: Build, test, package, deploy pipelines
Feature store
AUTOMATED PIPELINES
Data Engineering
ML Model Engineering
Model registry
ML Metadata store
CD Stage: ML Model serving
Trigger
Performance monitoring
ML Prediction service
@visenger

⑨ KNOW THE HIDDEN COSTS OF ML ENGINEERING

@visenger

10 BUILD RESPONSIBLE ML MODELS

be aware of ethical issues of AI

- DATA COLLECTION
- DATA REPRESENTATION
- ML MODEL STRUCTURE

- BIAS
- FAIRNESS
- PRIVACY
- INTERPRETABILITY

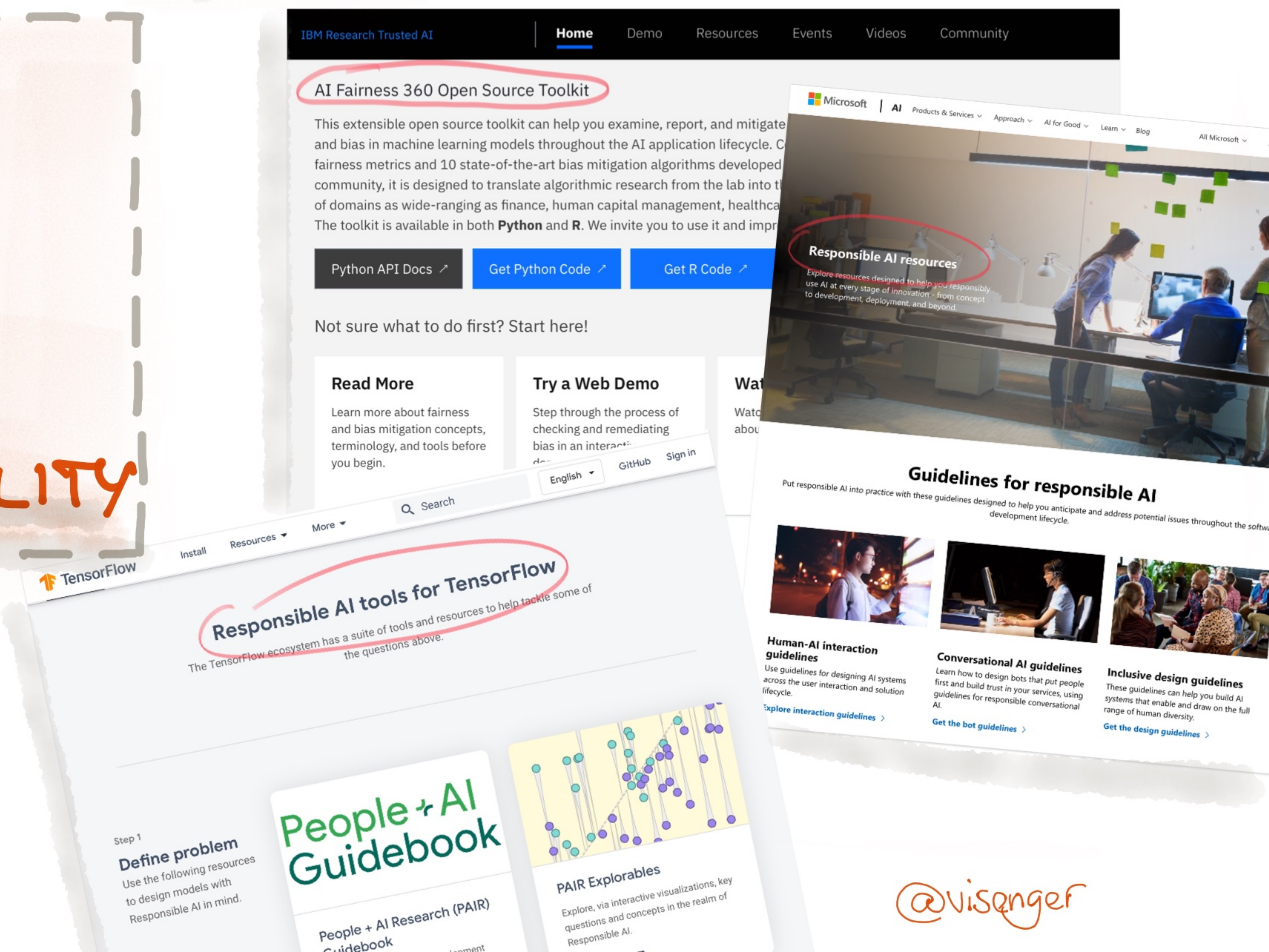
@visenger

10 BUILD RESPONSIBLE ML MODELS

ethical issues of AI

- BIAS
- FAIRNESS
- PRIVACY
- INTERPRETABILITY

tools & resources



@visenger

ML ENGINEERING HIERARCHY OF NEEDS

RESPONSIBLE AI

KNOW ML HIDDEN COSTS

DESIGN FOR ML

VERSIONING FOR ML

ML OPERATIONS

ML ALGORITMS SIMPLICITY

NO AI WITHOUT IA

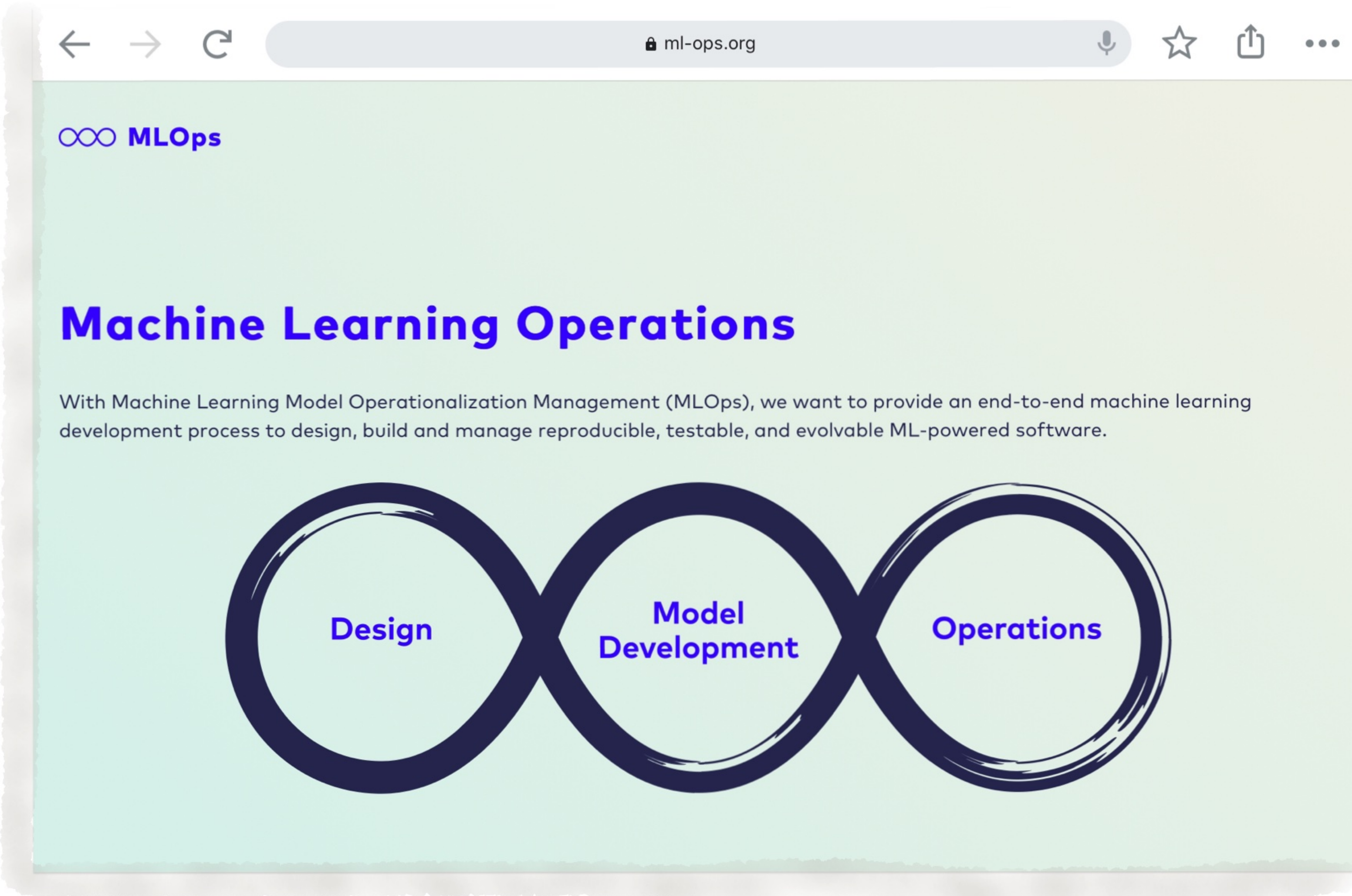
ML PROJECT MANAGEMENT

GET THE RIGHT TEAM/SKILLS

SOLVE ML PROBLEM

@visenger

MLOps @ INNOQ


ml-ops.org
MLOps
Machine Learning Operations
With Machine Learning Model Operationalization Management (MLOps), we want to provide an end-to-end machine learning development process to design, build and manage reproducible, testable, and evolvable ML-powered software.
Design
Model Development
Operations