SFU SIMON FRASER UNIVERSITY
SCHOOL OF INTERACTIVE ARTS & TECHNOLOGY

Interactive Visualization to Externalize, Explore, and Explain Trust in ML

Presenter: Brian Fisher

S. v. d. Elzen *et al.*, "The Flow of Trust: A Visualization Framework to Externalize, Explore, and Explain Trust in ML Applications," in *IEEE Computer Graphics and Applications*, vol. 43, no. 2, pp. 78-88, 1 March-April 2023, doi: 10.1109/MCG.2023.3237286.

INTEGRATED Science Lab

SFU SI
UN

Dagstuhl Seminar 22351

Interactive Visualization for Fostering Trust in ML

(Aug 28 – Sep 02, 2022)

SFU SIMON FRASER UNIVERSITY

My role: Psychology of Technology

- Ph.D Cognitive Psychology @ U Cal Santa Cruz
- PDA @ Institute for Robotics ^ Intelligent Systems @ Western U
- Research Prof. Cognitive Science @ Rutgers
- Inst. for Human Factors & Interface Technologies @SFU
- Media & Graphics Interdisciplinary Centre @ UBC
  - Faculty affiliations in Commerce, Computer Science, & Psychology
- School of Interactive Arts and Technology @ SFU
  - Creative design + engineering design + behavioural science
- IEEE Computer Society community: VIS GC, VEC, VGTC

Visual Analytics "The science of analytical reasoning facilitated by interactive visual interfaces"

SFU

“From Visualization to Visually-Enabled Reasoning”

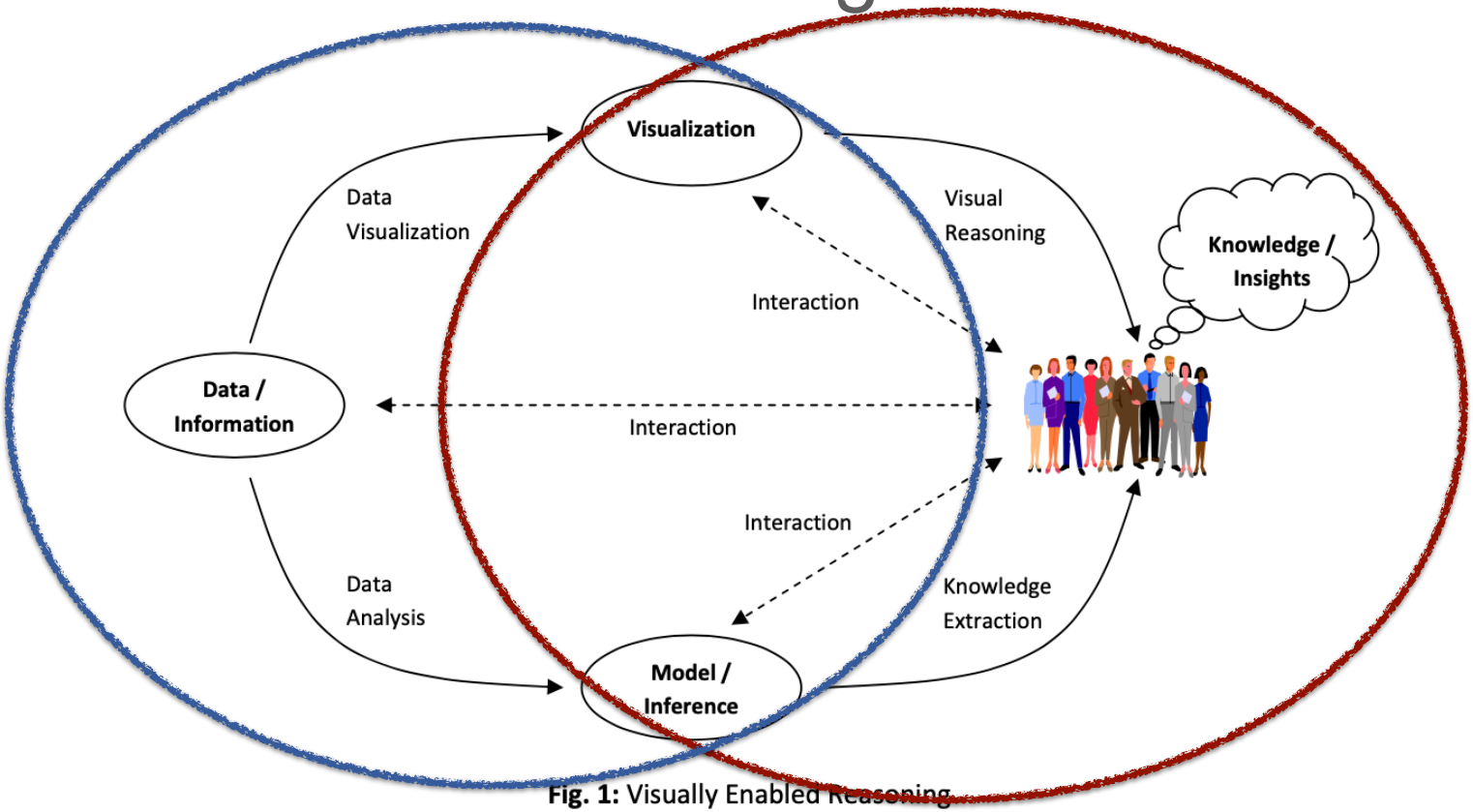


Fig. 1: Visually Enabled Reasoning

Meyer J., Thomas, J., Diehl, S., Fisher, B., Keim, D., Laidlaw, D. Miksch S., Mueller, K. Ribarsky, W., Preim, B., & Ynnerman, A. (2010) From Visualization to Visually Enabled Reasoning. In “Scientific Visualization: Advanced Concepts”. vol. 1 pp. 227-245. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany. I978-3-939897-19-4

SFU

Cognitive science grounding

SFU

Thinking about trust

"Firm belief in the reliability, truth, or ability of someone or something; confidence or faith in a person or thing, or in an attribute of a person or thing"

Trust models

- Trust in computation
- Trust based on experience
- Trust in the developer
- Trust by expert confirmation
- Trust in the agent itself ?

Dennett Stances

- Physical: frame actions as predicted by structures
- Designed: frame actions as leading to design objectives
- Intentional: frame behaviours as produced by a cognitive agent using Theory of Mind (module?)

EUROVIS Leipzig 2023

25th EG Conference on Visualization
Leipzig, Germany 12 - 16 June 2023

The Flow of Trust: A Visualization Framework to Externalize, Explore & Explain Trust in ML Applications

Results of Dagstuhl Seminar on *Interactive Visualization for Fostering Trust in ML (seminar 22351)*

Stef van den Elzen

TU/e Eindhoven University of Technology Visualization

Gennady Andrienko

Fraunhofer IAIS

Natalia Andrienko

City University London

Brian D. Fisher

SFU

Rafael M. Martins

Linnæus University Sweden

Jaakko Peltonen

Tampere University

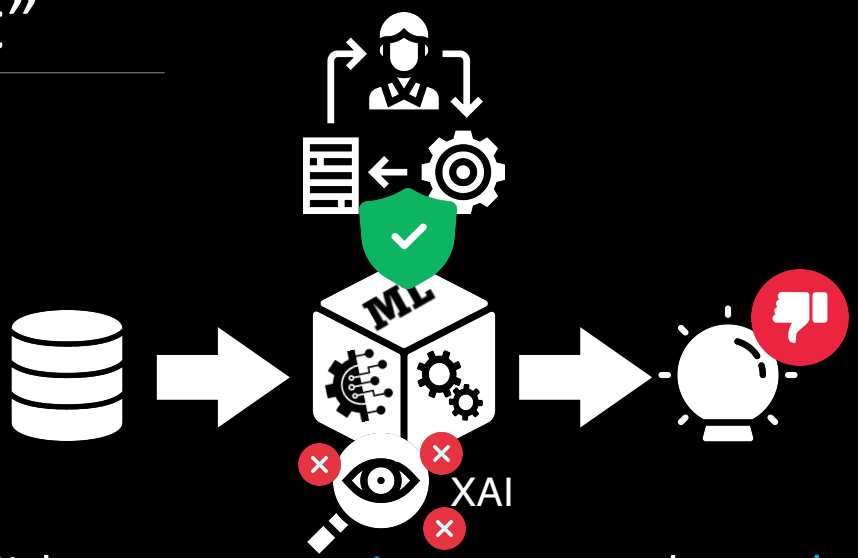
Alexandru C. Telea

Utrecht University

Michel Verleysen

UCLouvain

“The Flow of Trust”
Motivation



ML becomes more important and complex

Explainable AI (XAI) helps to understand the models and/or output

Do not directly cover building **trust** in the model.

Motivation

Trust in ML applications is
an implicit process that
takes place in the user's mind.

VA & ML applications lack an interface for
expressing trust and/or distrust.

No method of feedback or communication
of trust that can be acted upon.

Motivation

Endert et al., 2017

Chatzimparmpas et al., 2020

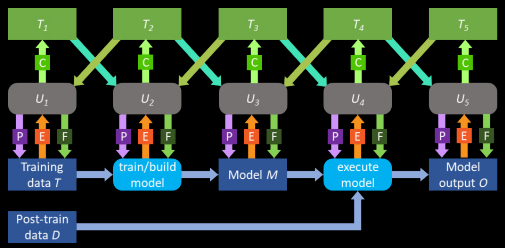
Sperrle et al., 2021

Trust one of the most important goals.

Do not discuss how to directly achieve that in a concrete manner.

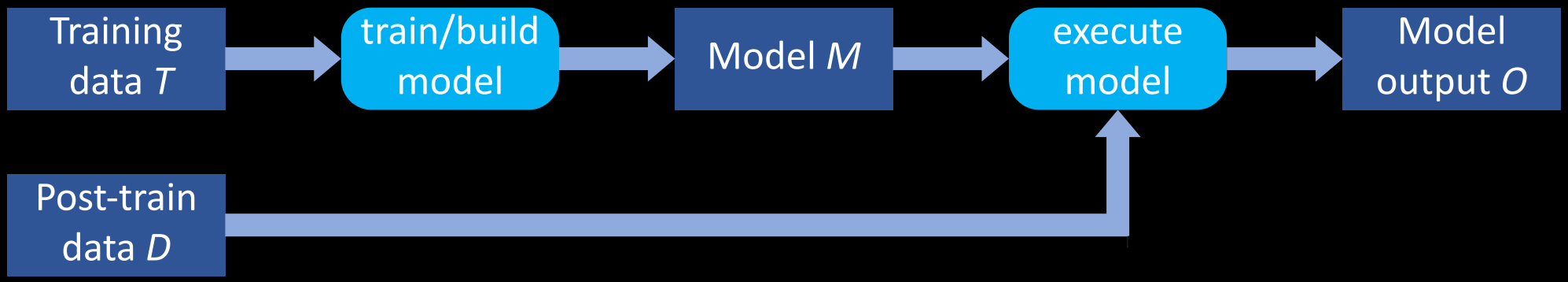
Motivation

Trust as first-class citizen.

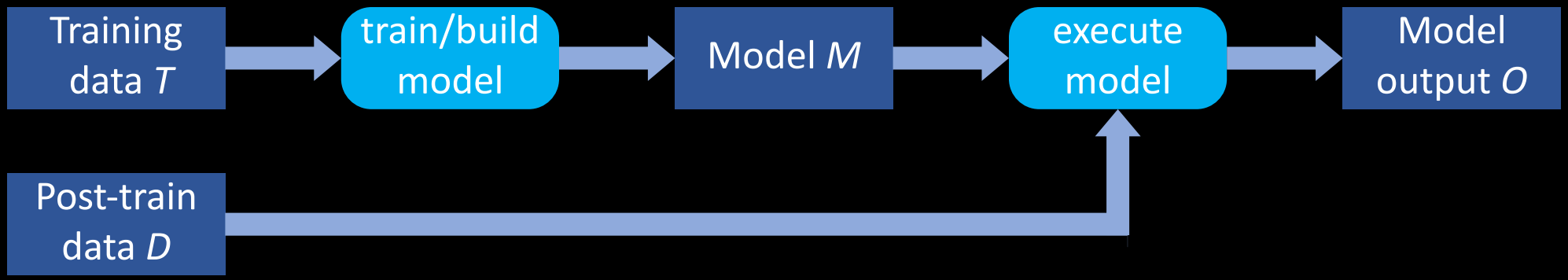


Conceptual framework that captures the flow of trust.

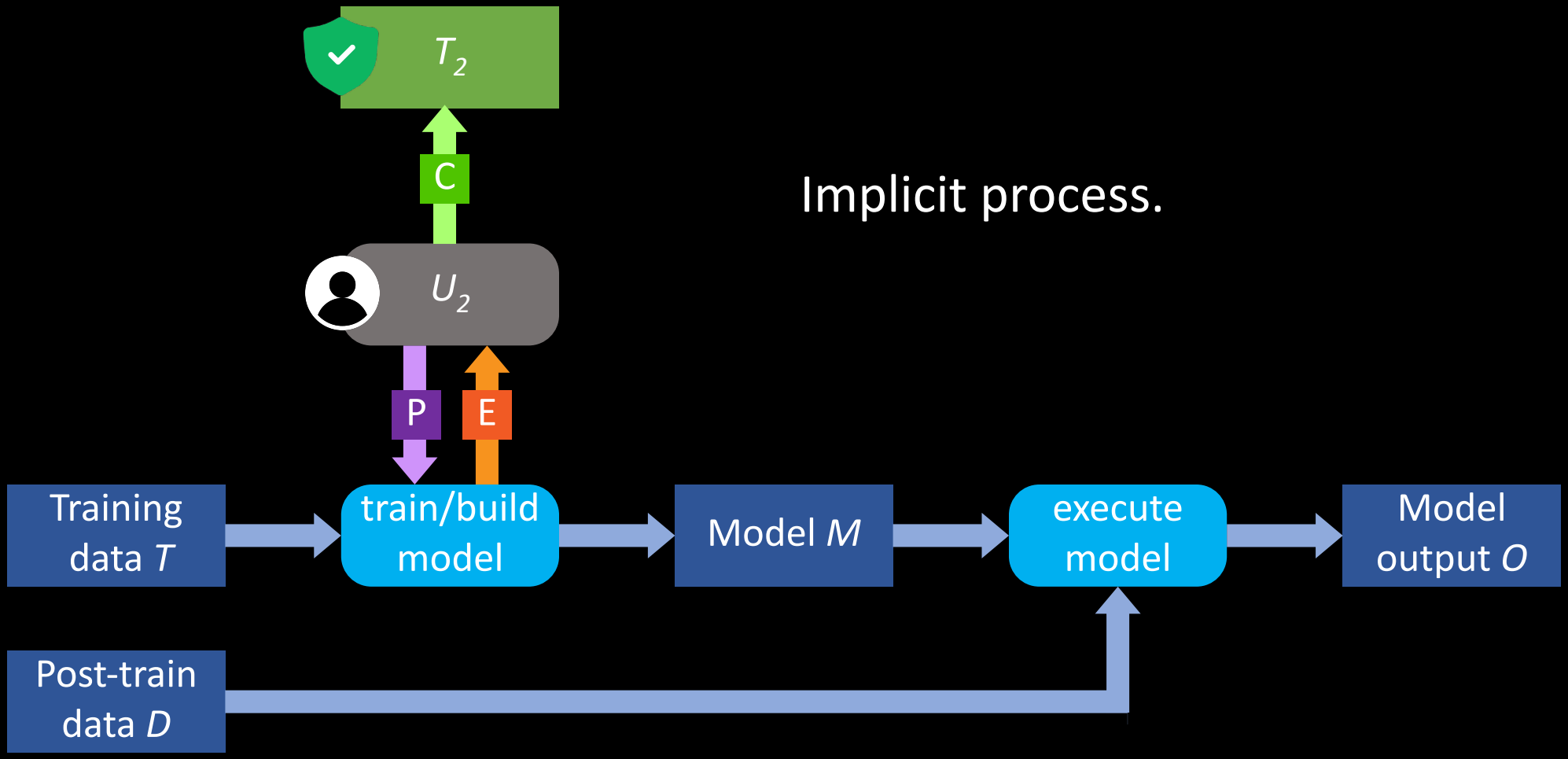
ML pipeline


Training data *T*
train/build model
Model *M*
execute model
Model output *O*
Post-train data *D*

ML pipeline

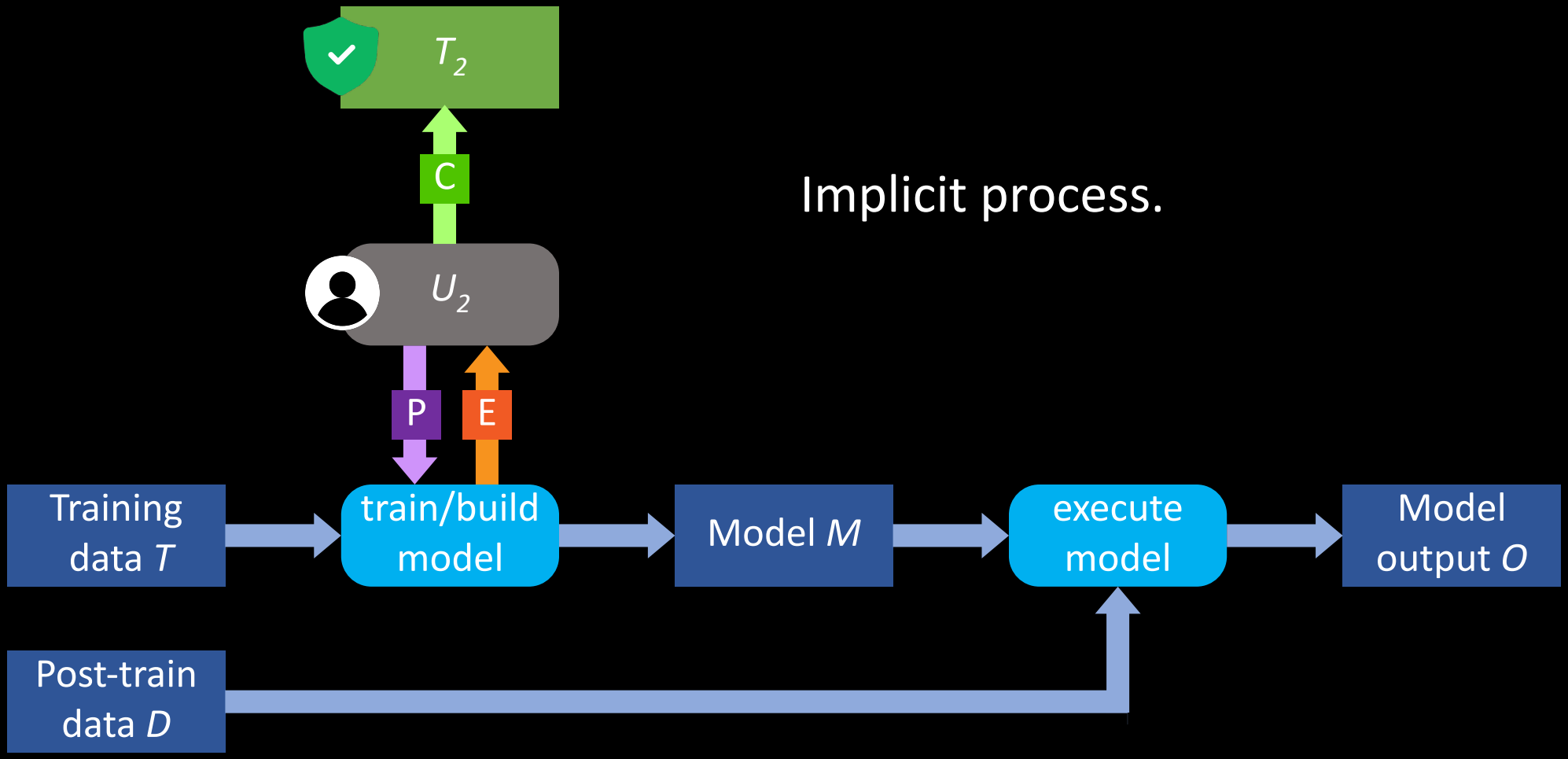

Training data *T*
train/build model
Model *M*
execute model
Model output *O*
Post-train data *D*

Trust as first-class citizen



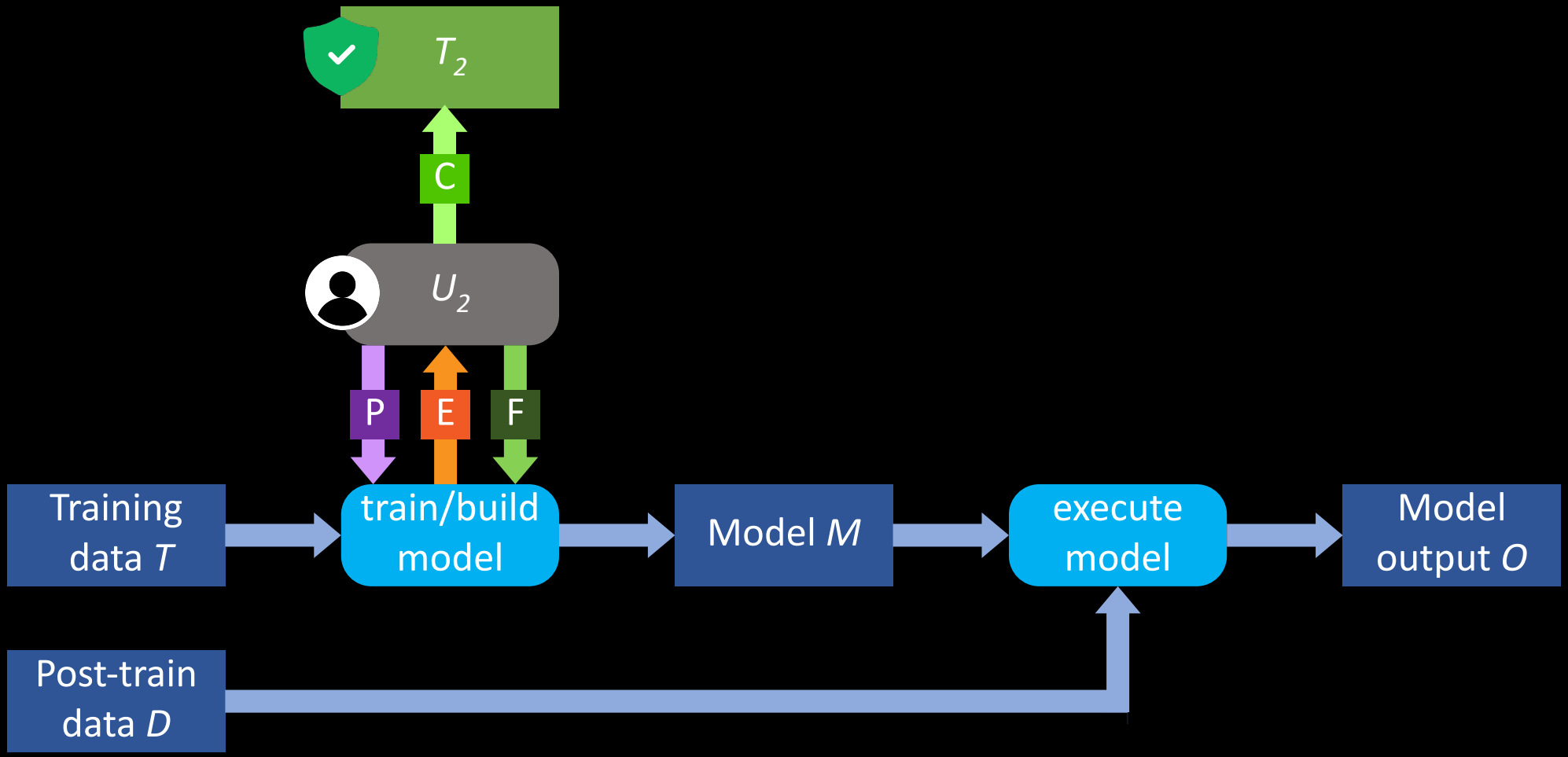
Implicit process.

Trust as first-class citizen

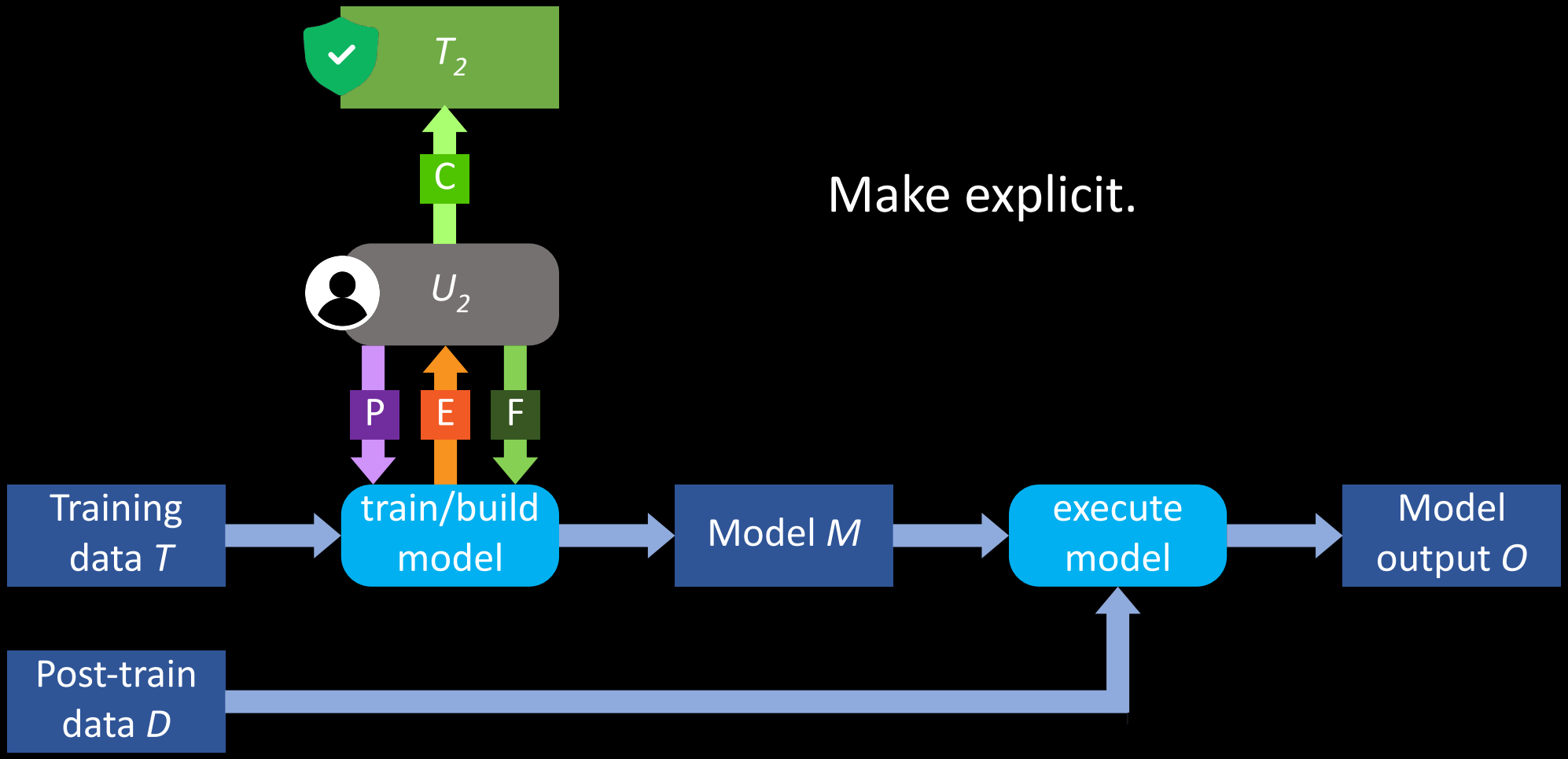


Implicit process.

Trust as first-class citizen

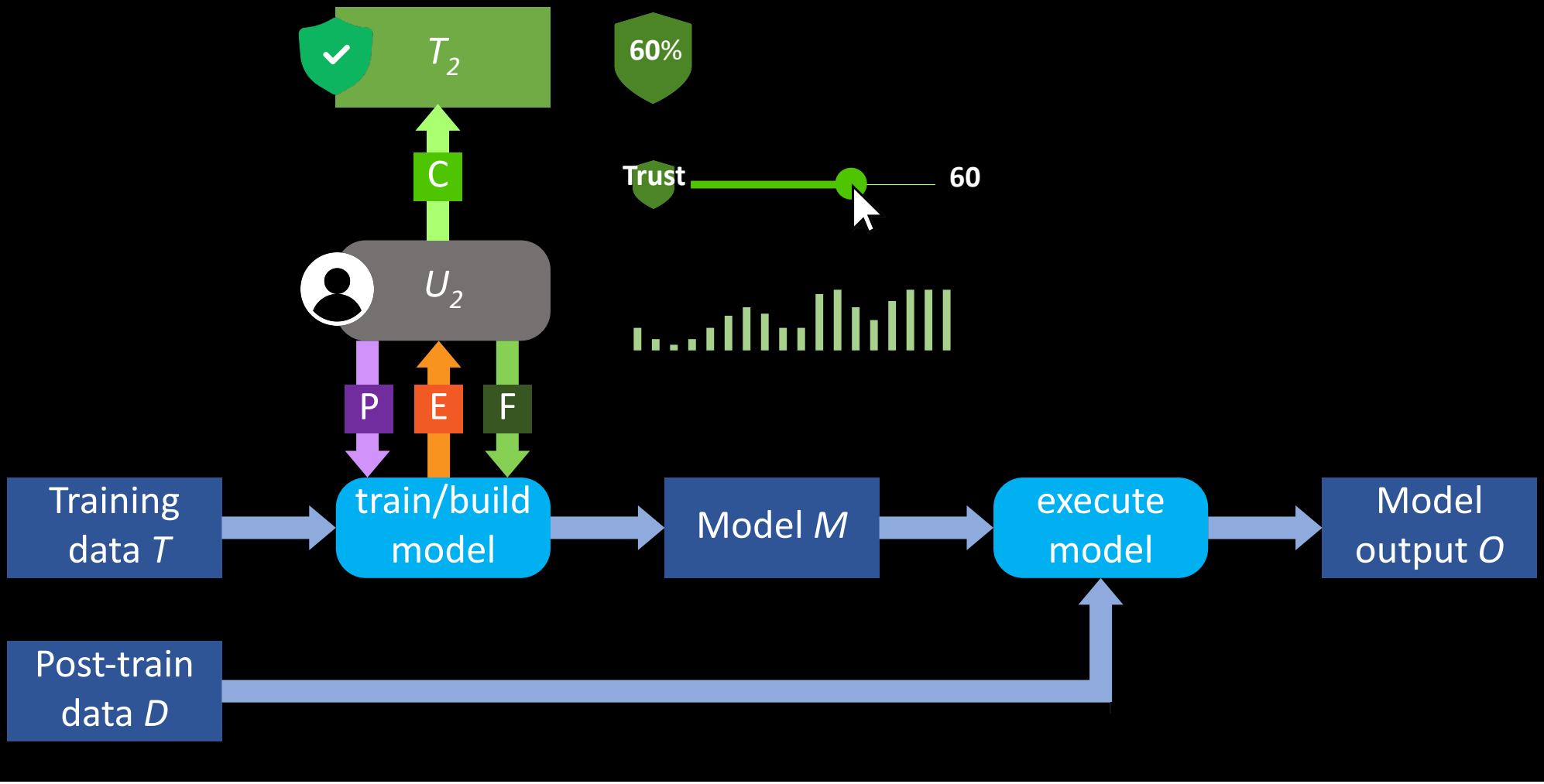


Trust as first-class citizen

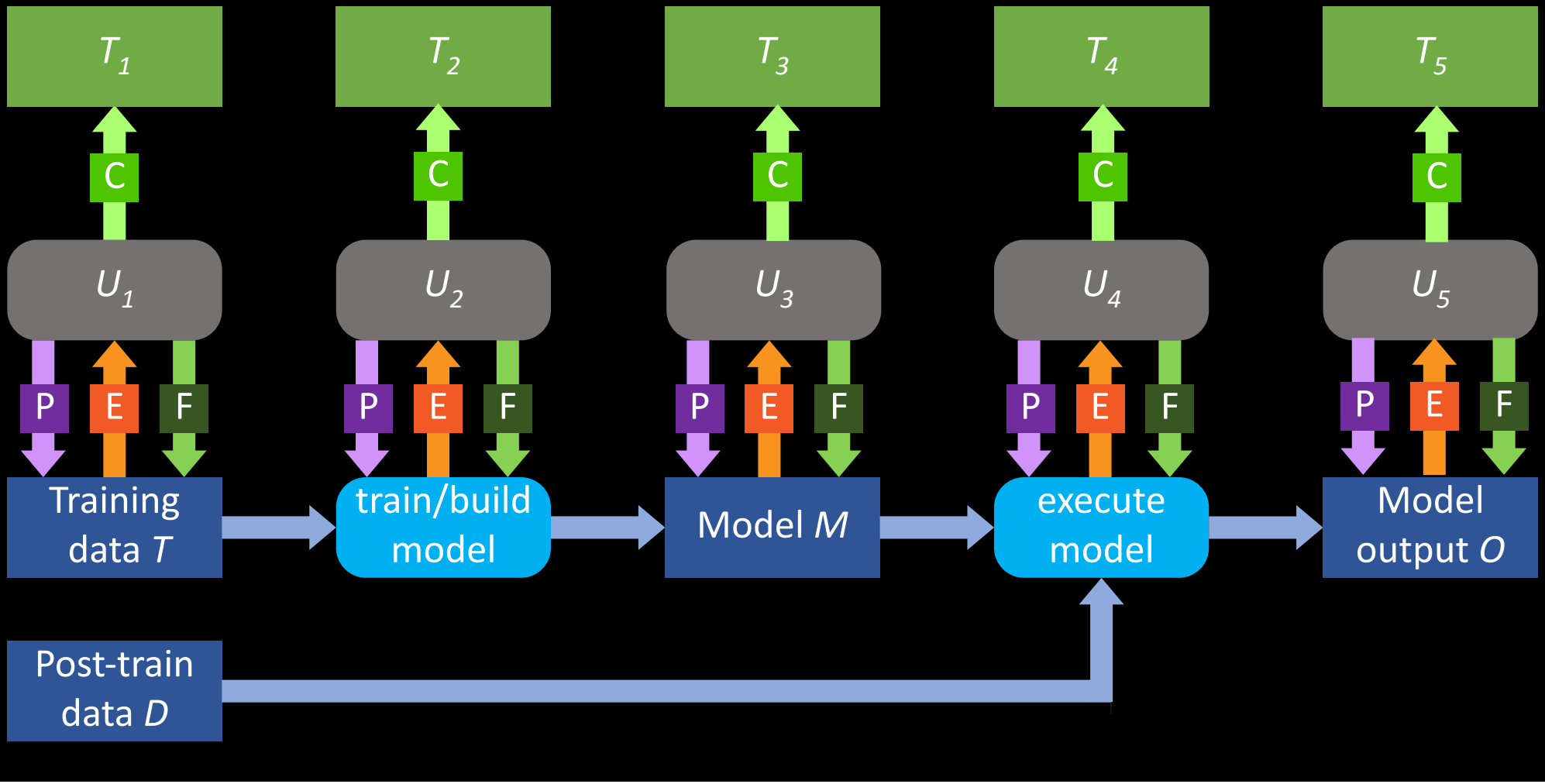


Make explicit.

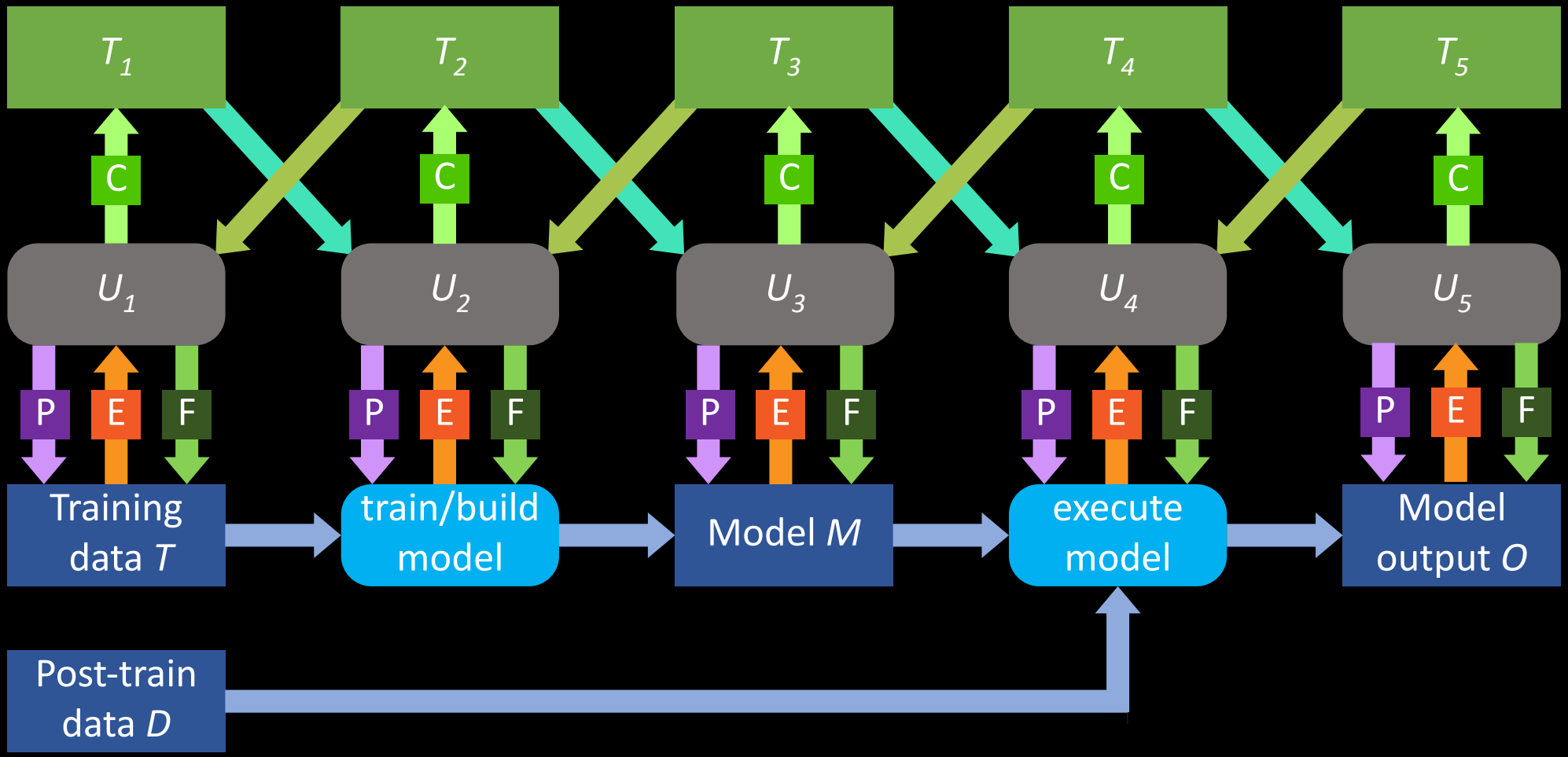
Trust as first-class citizen



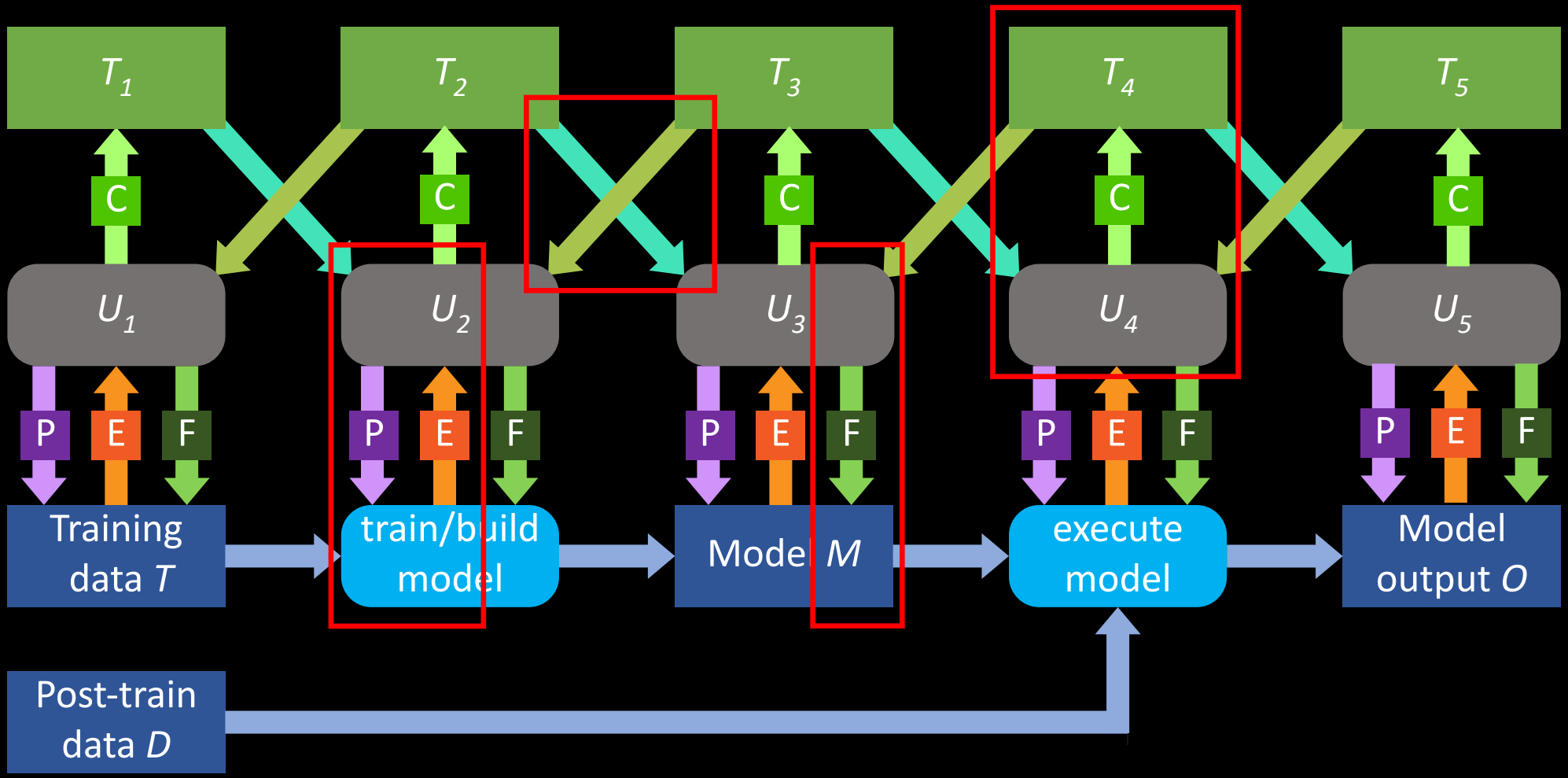
Flow of trust



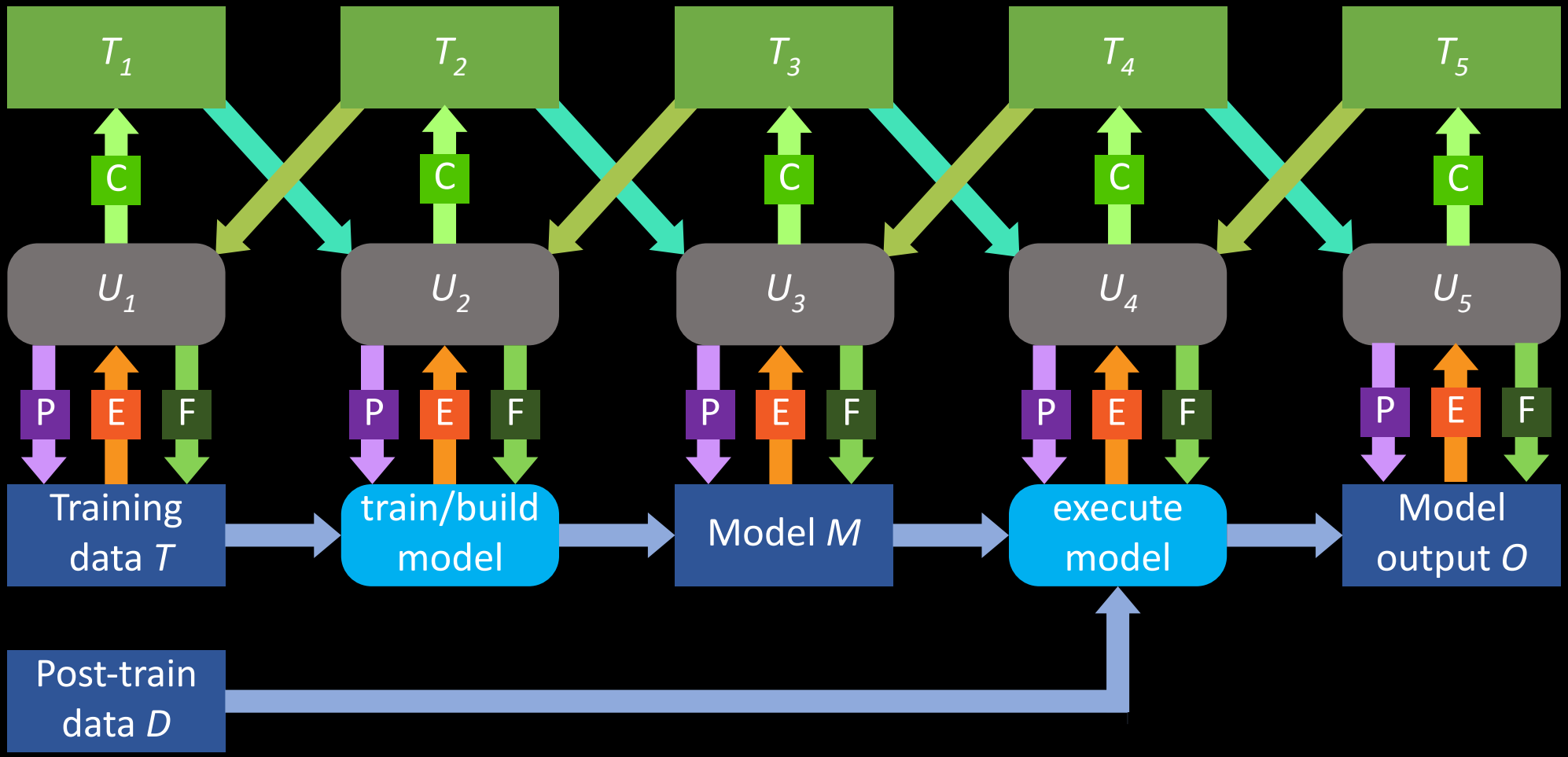
Flow of trust



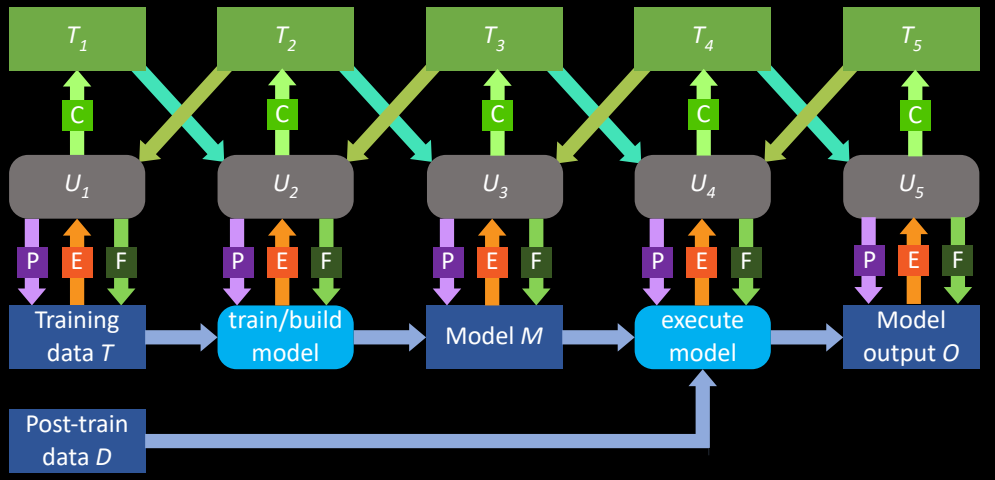
Role of interactive visualization



Flow of trust



Flow of trust



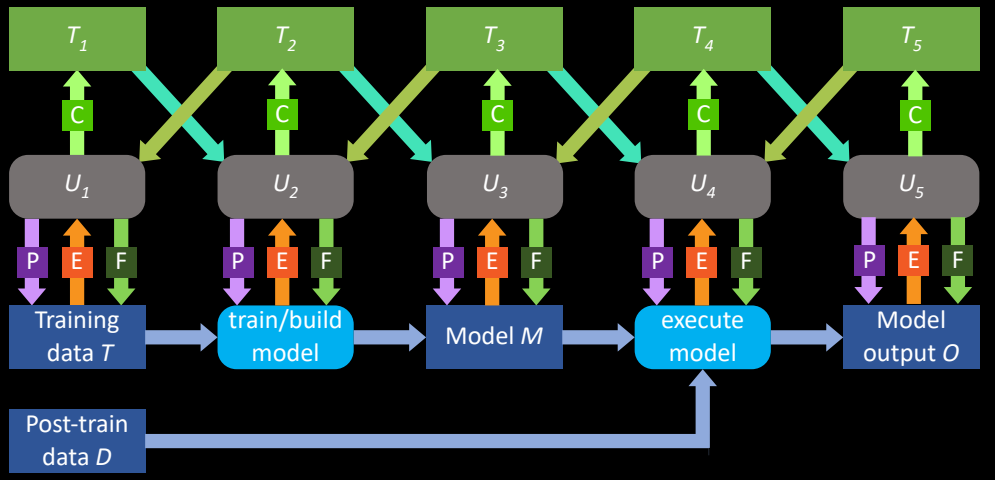
Intended use of the framework

Define and guide new research area in VA

Trust externalization

- Different for each object

How represent & combine trust?

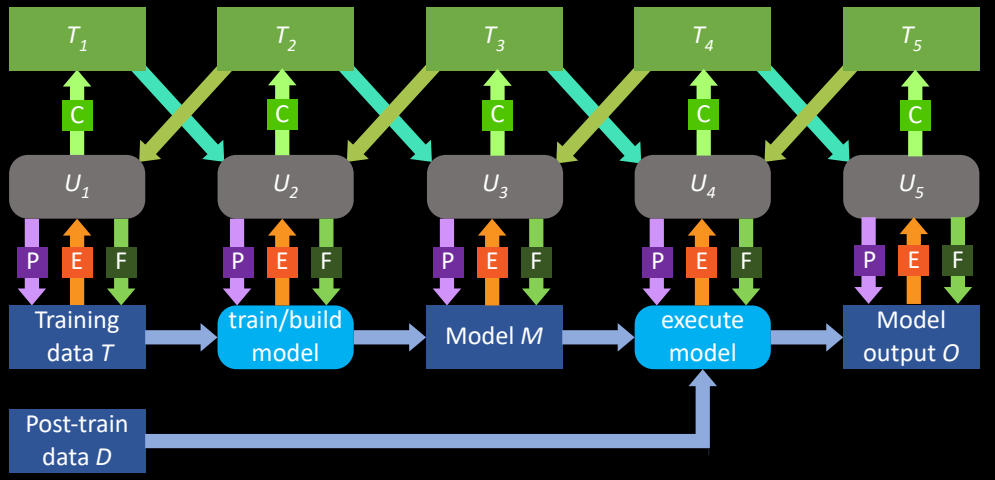


Trust expression and communication

Open areas for research

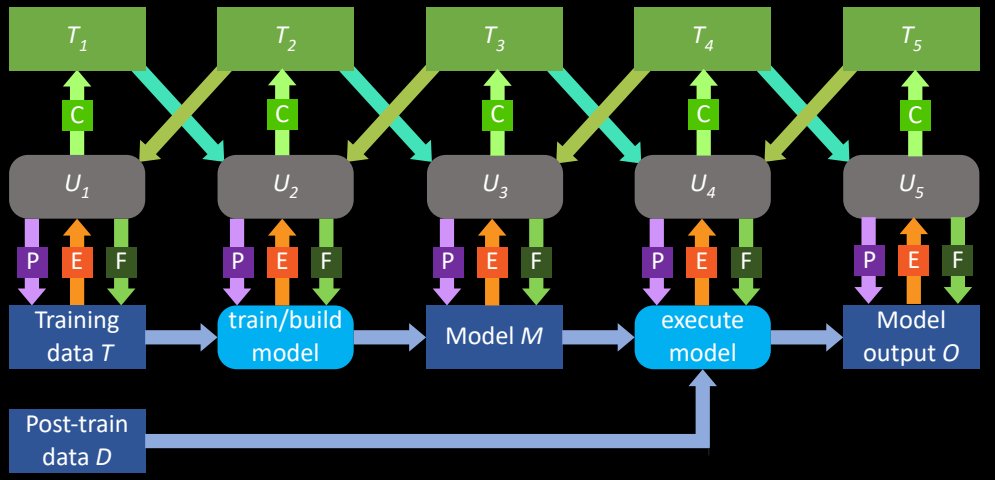
Open areas for research

1. Trust objects
   - Taxonomy
   - Trust issues
   - Possible reasons for (mis)trust



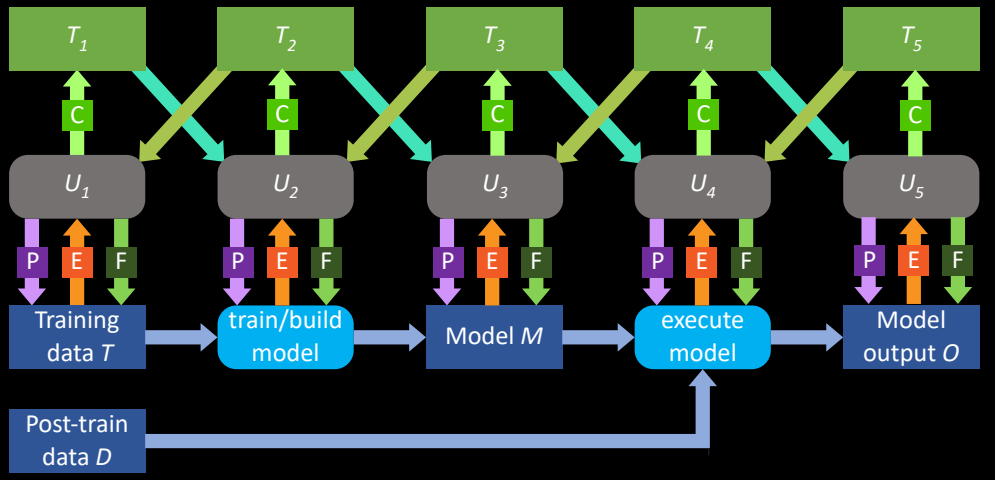
Open areas for research

2. Formalisms
   - Represent trust in machine readable form.



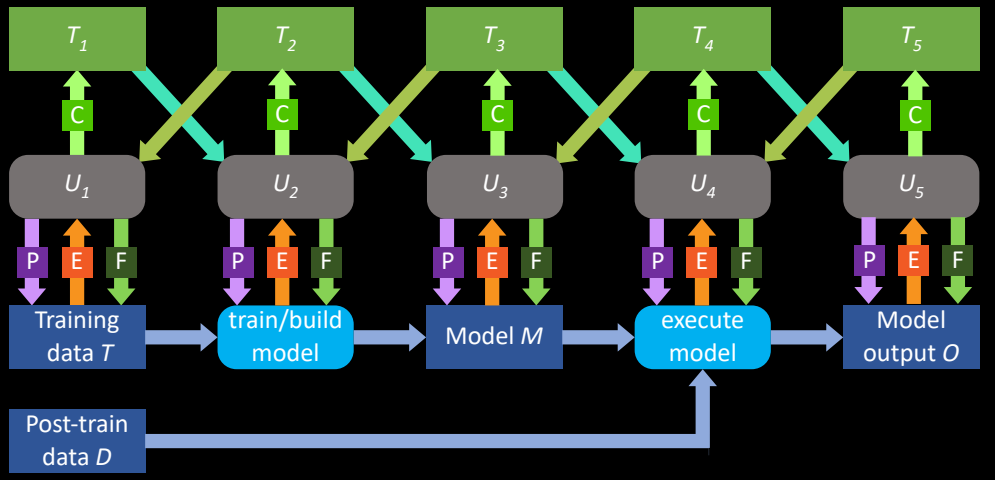
Open areas for research

3. Expression
   - Ways for users to express their state of trust by interacting with a computer system.



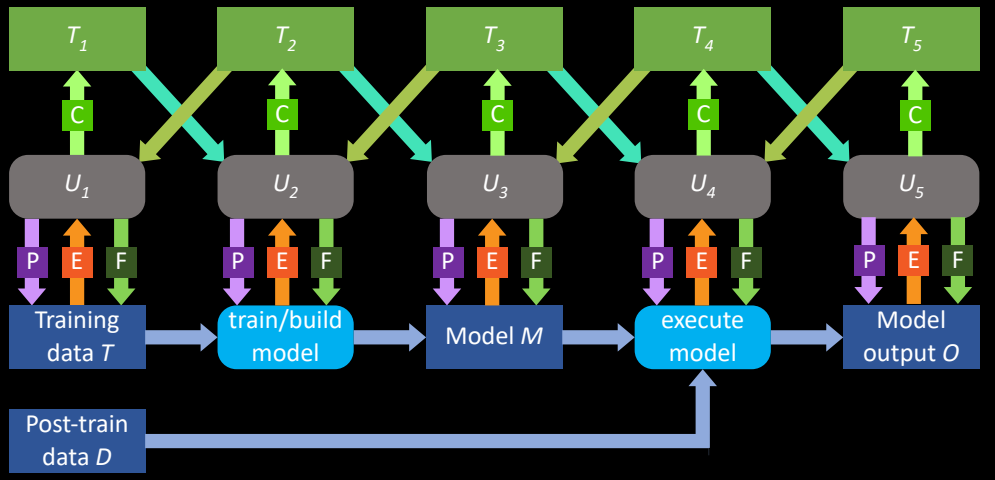
Open areas for research

4. Flow of trust
   - Ways to explore and develop trust over all stages of a ML pipeline using visual interactive techniques.



Open areas for research

5. Guidance
   - Ways to facilitate users' expression and communication of the state of trust using visual interactive techniques.

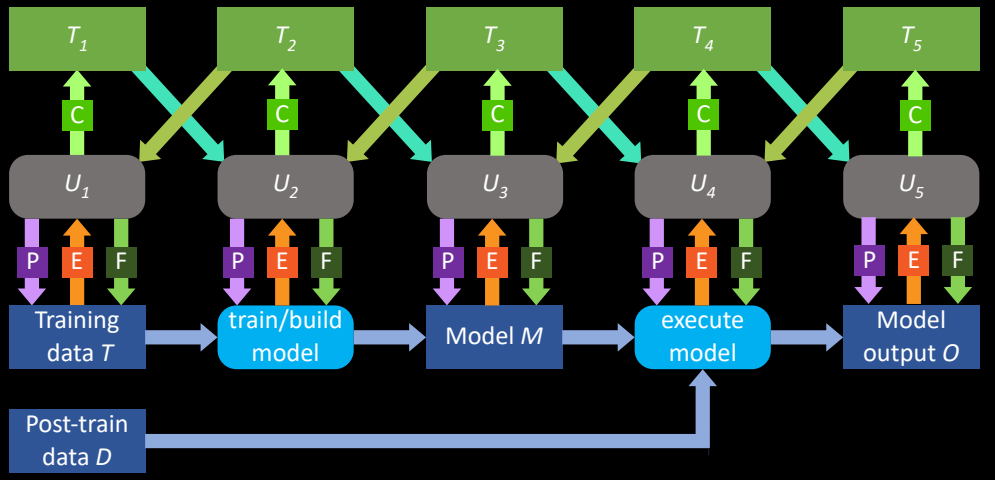


Conclusions

Trust in ML/VA applications is
- an implicit process
- taking place in the user's mind.

No method of feedback or communication of trust that can be acted upon.

Our framework:
- Instrumental in developing interactive visualizations to help users build and communicate trust.
- Support the flow of trust *within* and *between* stages.



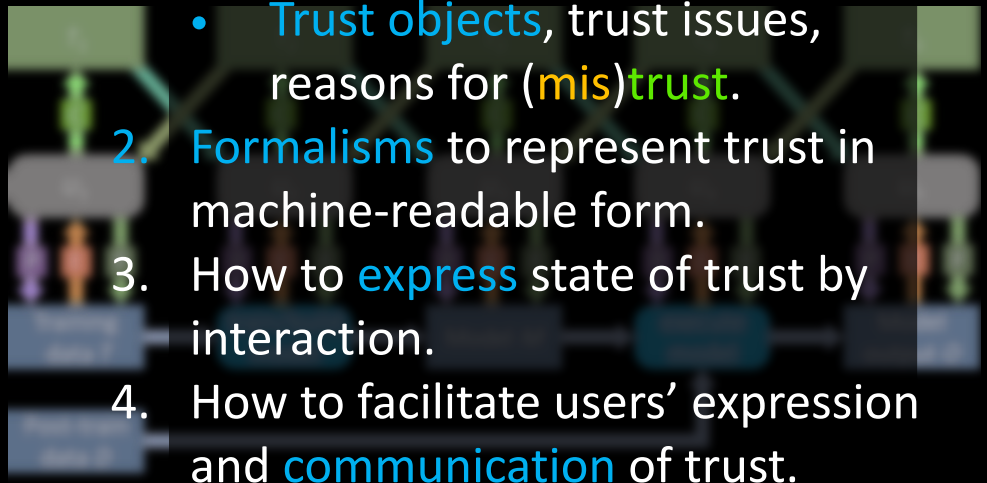
Conclusions

Trust in ML/VA applications is
- an implicit process
- taking place in the user's mind.

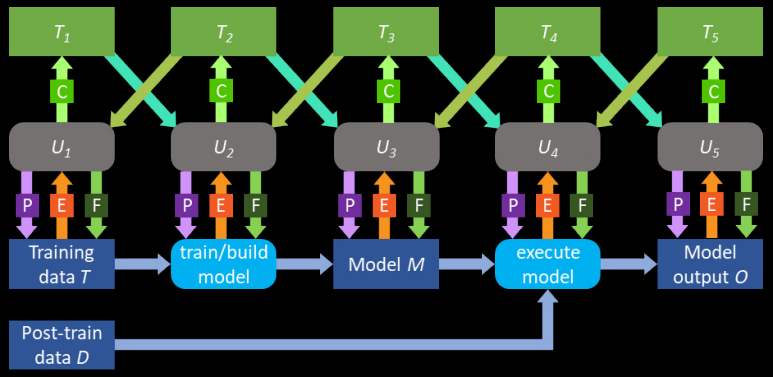
No method of feedback or communication of trust that can be acted upon.

Our framework:
- Instrumental in developing interactive visualizations to help users build and communicate trust.
- Support the flow of trust *within* and *between* stages.

Research questions and directions:
1. Typology/taxonomy of
   - Trust objects, trust issues, reasons for (mis)trust.
2. Formalisms to represent trust in machine-readable form.
3. How to express state of trust by interaction.
4. How to facilitate users' expression and communication of trust.
5. Visual interactive techniques for representation and exploration of trust.



The Flow of Trust: A Visualization Framework to Externalize, Explore & Explain Trust in ML Applications



TU/e Eindhoven University of Technology Visualization

Fraunhofer IAIS

Linnæus University Sweden

Tampere University


Stef van den Elzen


Gennady Andrienko


Natalia Andrienko


Brian D. Fisher


Rafael M. Martins


Jaakko Peltonen


Alexandru C. Telea


Michel Verleysen

City University London

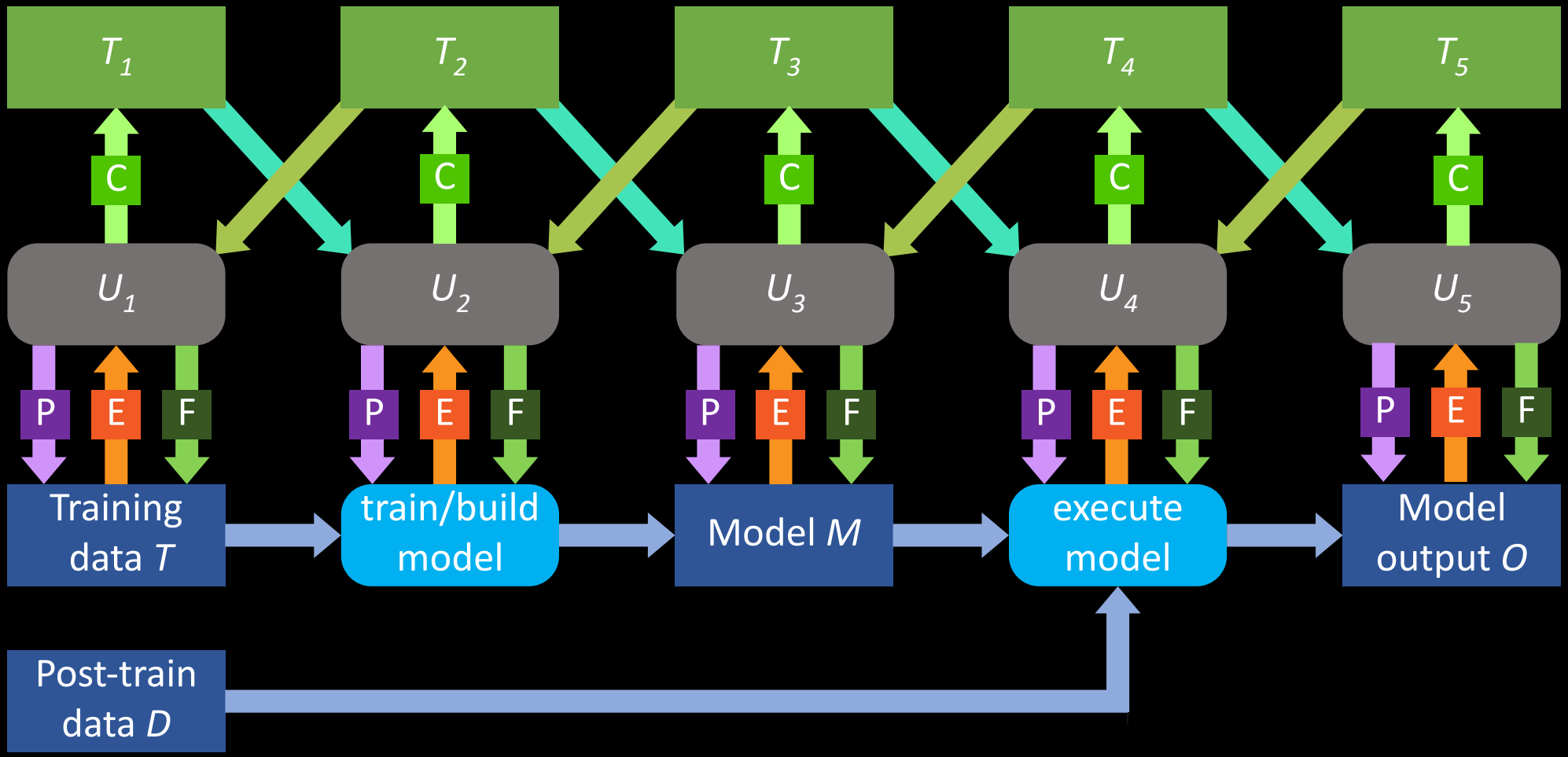
SFU

Utrecht University

UCLouvain

S. v. d. Elzen *et al*., "The Flow of Trust: A Visualization Framework to Externalize, Explore, and Explain Trust in ML Applications," in *IEEE Computer Graphics and Applications*, vol. 43, no. 2, pp. 78-88, 1 March-April 2023, doi: 10.1109/MCG.2023.3237286.

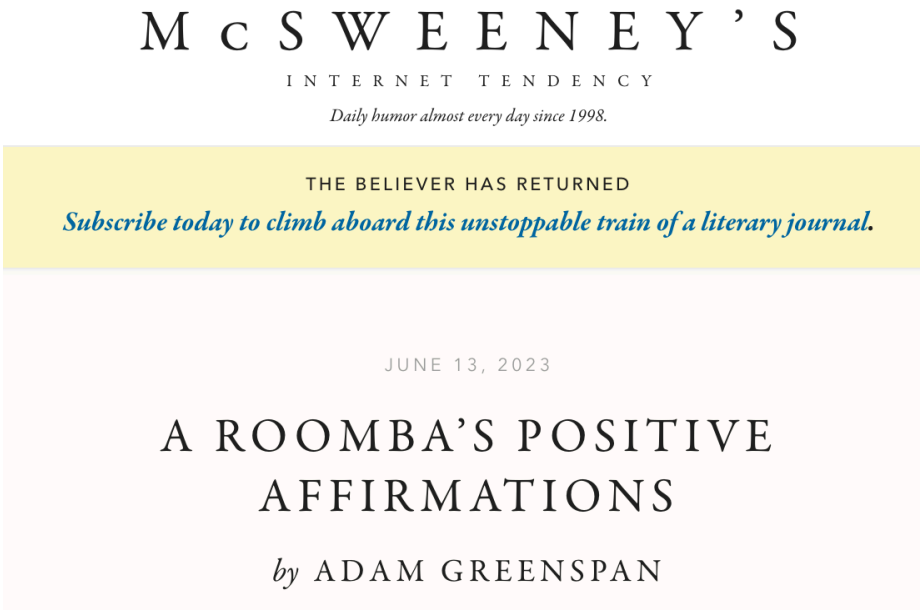
Flow of trust



SFU

Intentional Stance

Intentional: frame behaviours as produced by a cognitive agent using Theory of Mind (module?)



I am free of the boxes people put me in.

I am plugged in.

I am fully charged.

I am unstoppable.

I am running into a chair.

I am running into a chair.

….

I am the best at running into a chair.

I use obstacles to learn and grow.

I am this house's cleanest pet.

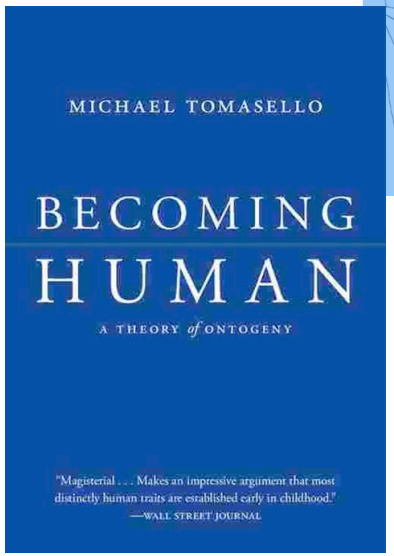
I am filled with love, happiness, and cat hair.

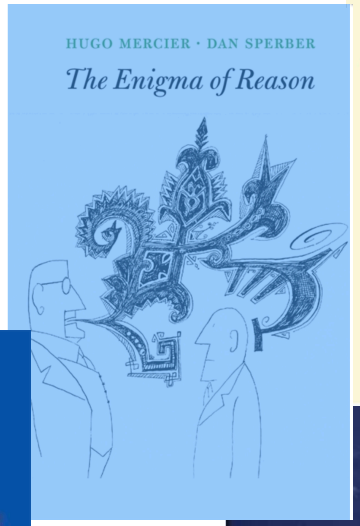
I am surrounded by love, happiness, and chairs.

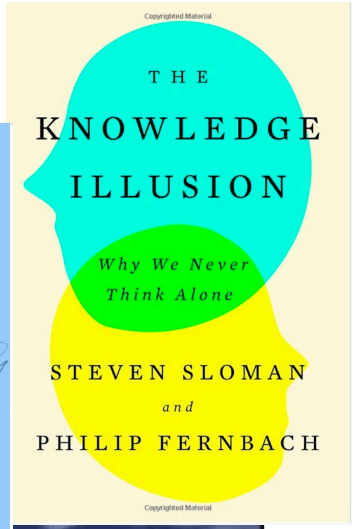
…..

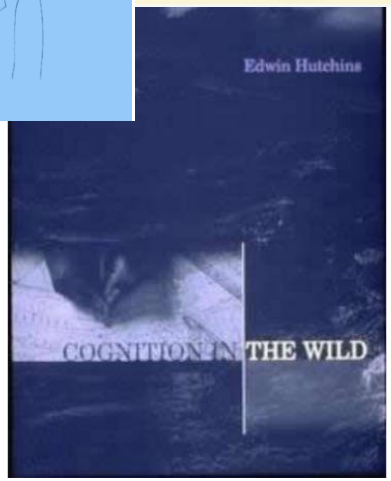
SFU

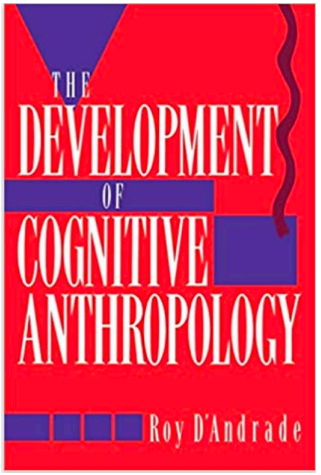
Understanding intelligent agents

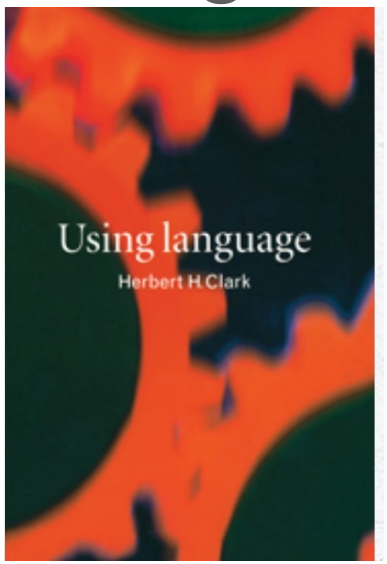


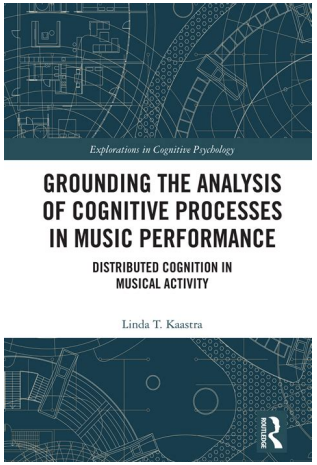






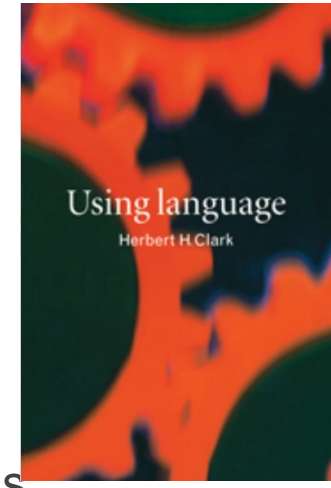




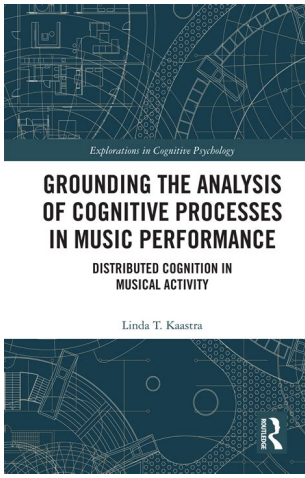


SFU

Joint Activity Theory (Clark)



- H. H. Clark's theory
  - Collaboration as a coordination task
  - Based on "common ground"
    - Processes by which common ground is developed through joint actions
    - Repair methods when coordination fails
- Extend Clark to focus on technology as integral to communication and collaborative analysis
- Can JAT be used to help design human-AI collaboration?



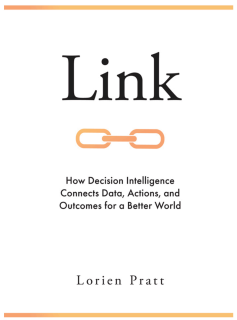
SFU

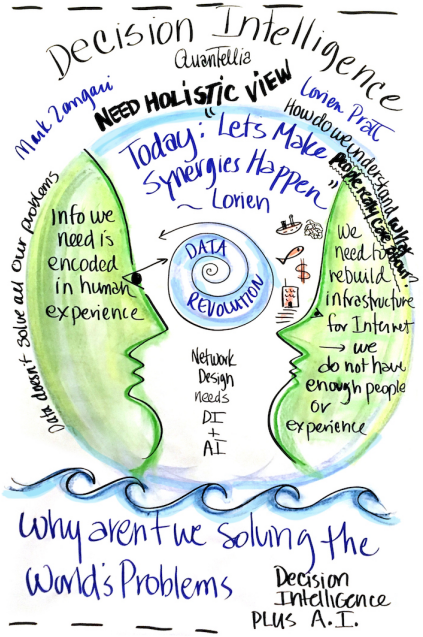
Decision Intelligence

SFU

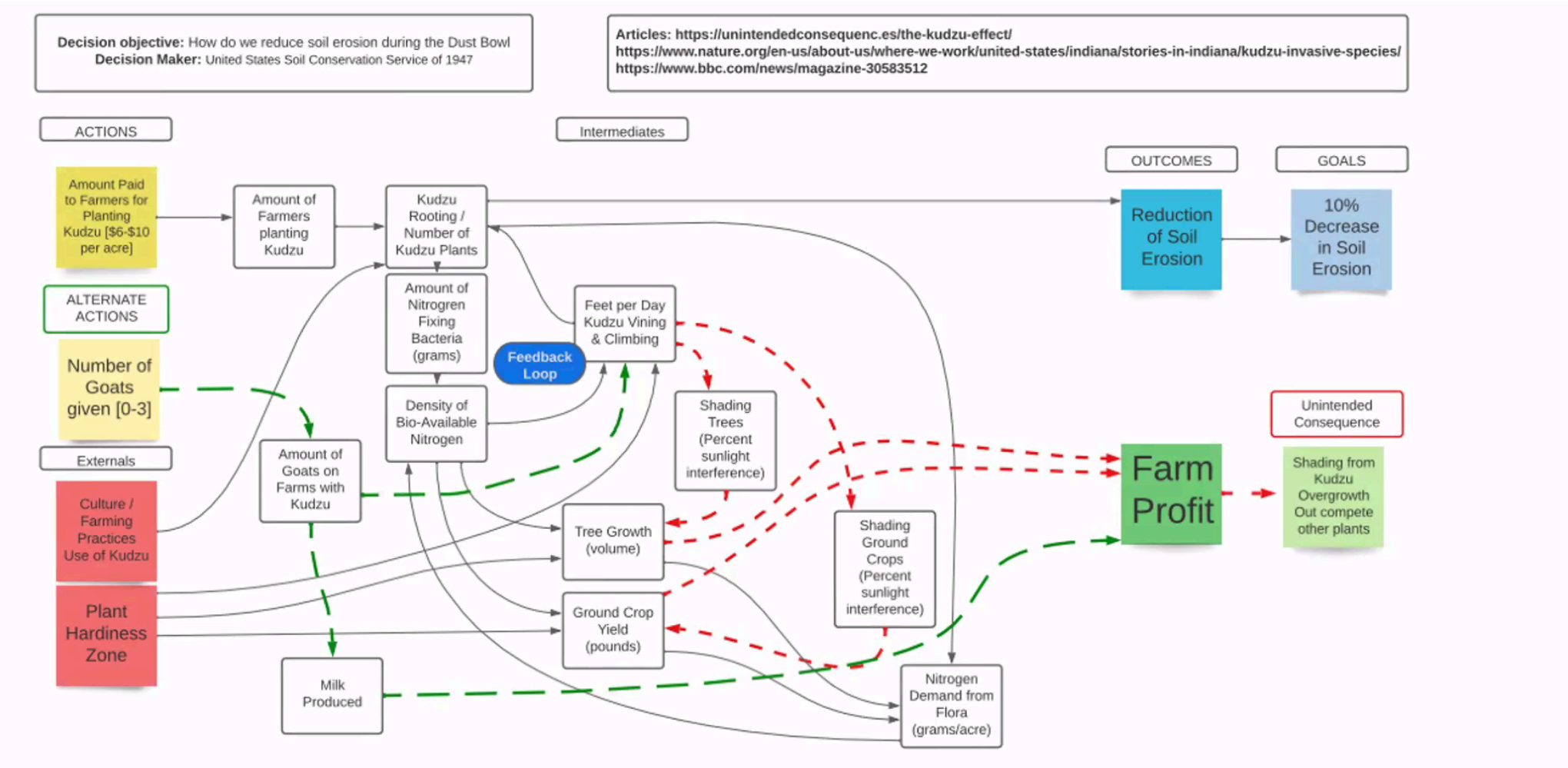
Decision pipeline

- CDD business decisions to likely outcomes
  - Elicit actions, externals, intermediates, outcomes from stakeholders
- Building CDD
  - Causality
  - Modal & Hybrid logics
  - Individual & group reasoning
  - Structured processes (e.g. Delphi)
- Build sim with AI calculation of intermediates





CDD example (with goats!)



SFU

Trust & CDD models (Pratt)

- Model creation process includes diverse stakeholders
- Model is transparent and readable by nontechnical people
- Model captures chains of reasoning that may lead to unintended consequences
- Model includes not just models built from (obscure, hard to understand or trace provenance) data, but also human knowledge, which is more explainable
- Model supports curation, review, and update one link at a time (to make it more tractable) as new knowledge becomes available
- Model supports inclusion of intangible factors like discomfort, cultural values, and more (because it's visual that expands the cognitive capacity of modelers)
- Model surfaces hidden agendas, exposes deception
- Model helps to overcome short-term thinking
- Model can avoid moral hazard by better balancing responsibility and authority