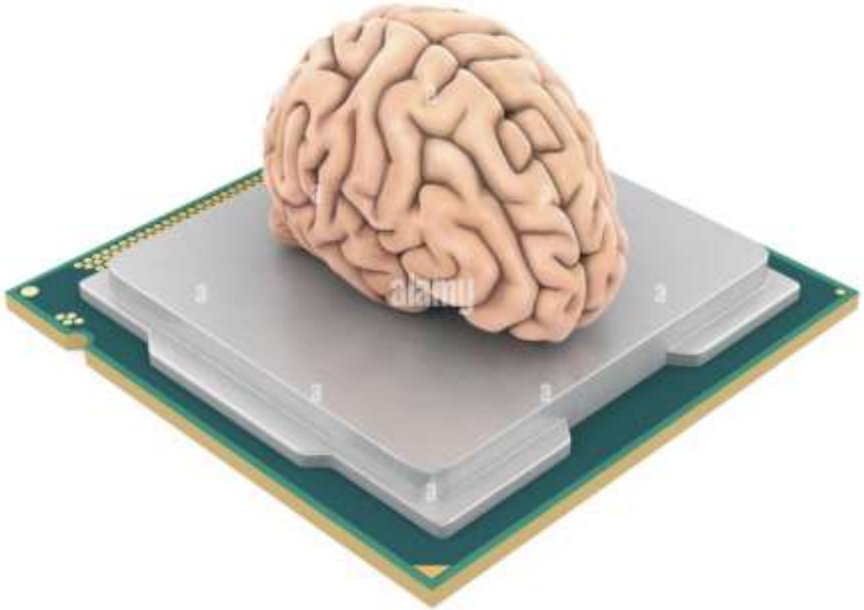


Human-AI interaction: Co-Pilots, Trainers, and Augmentation

Mark Chignell

Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

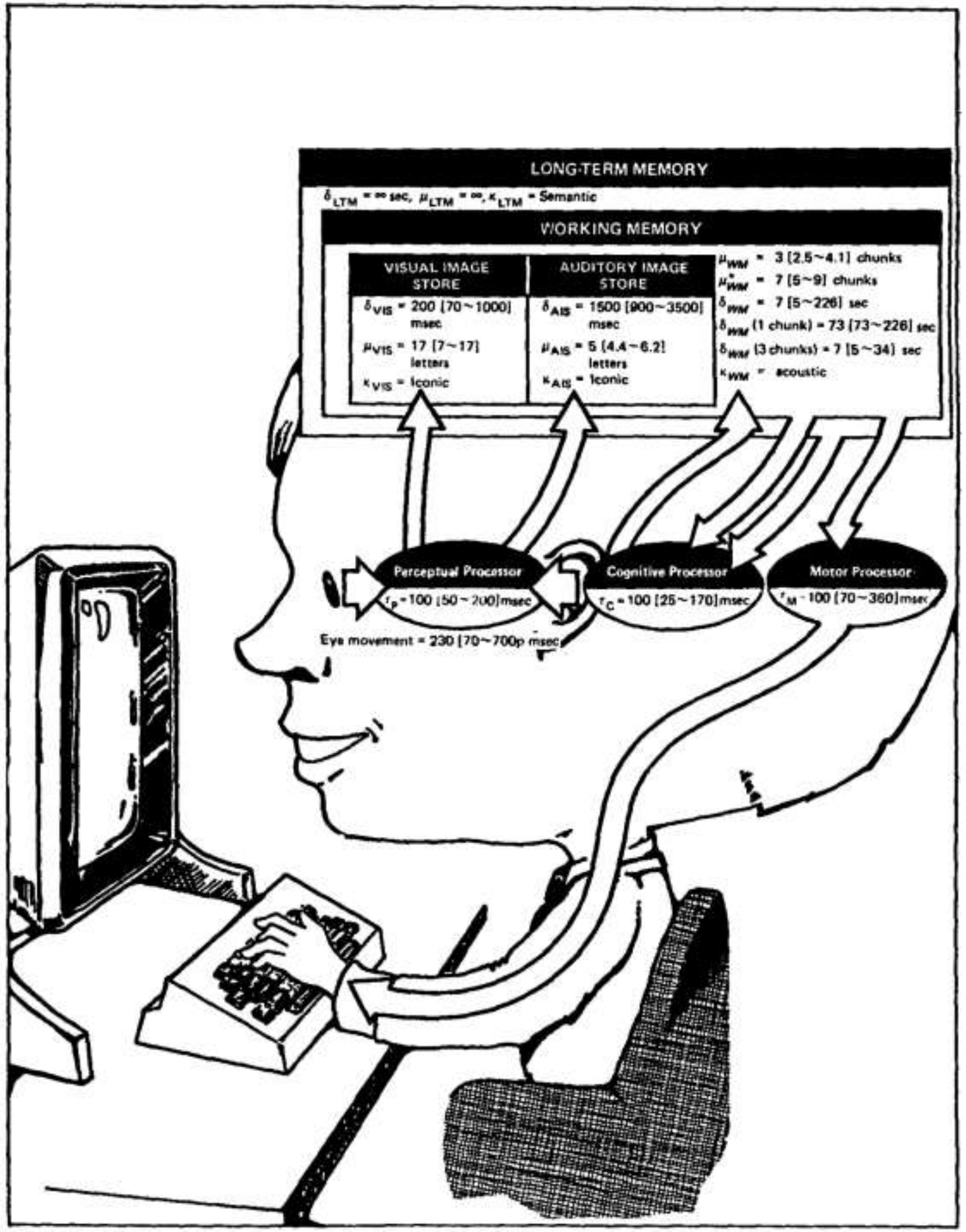
Interactive MediaLab

Overview

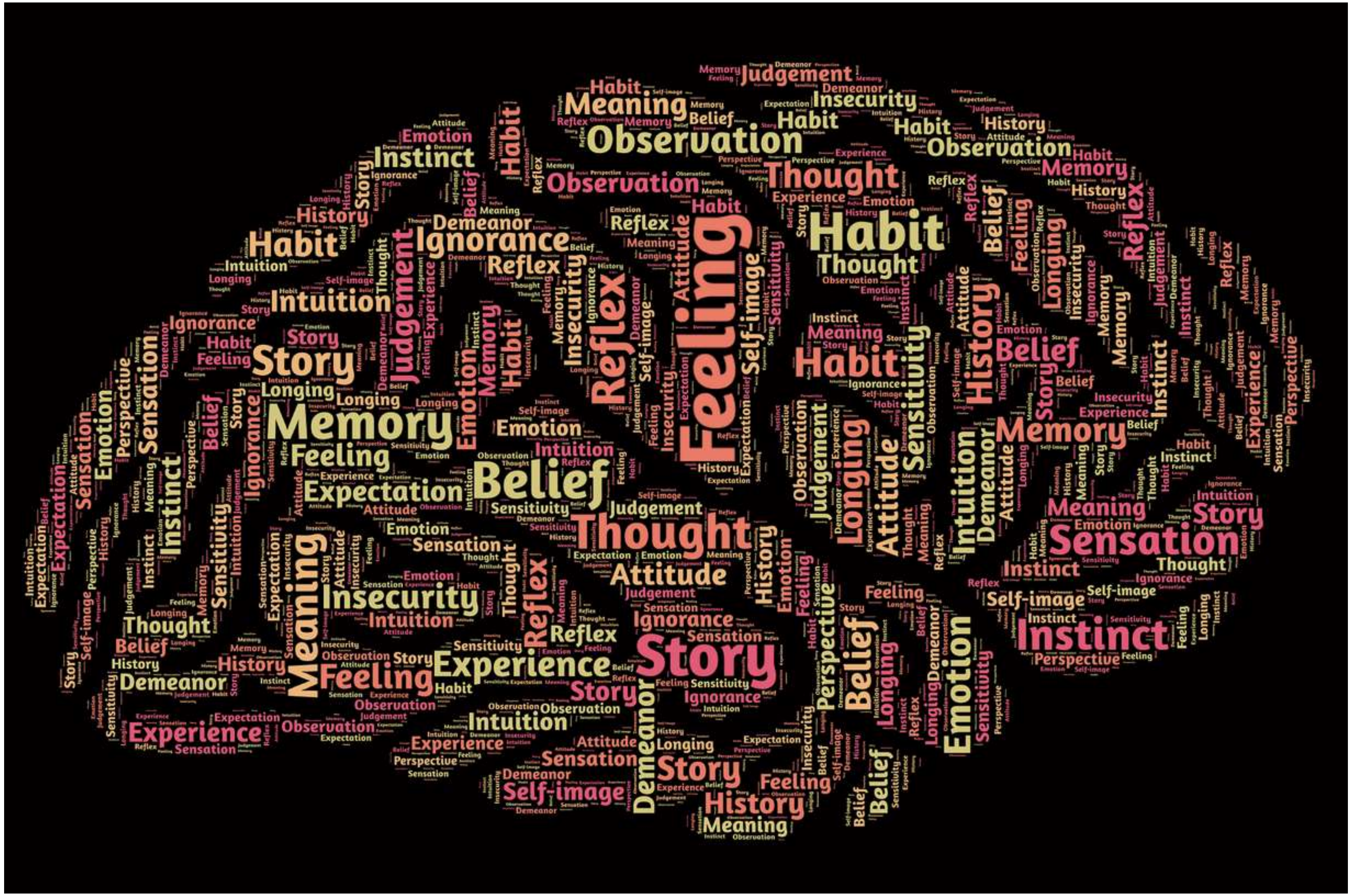
- The Human Brain
- The AI Challenge
    - Co-Pilots, Trainers and Augmentation
    - Augmentation of Humans vs. Augmentation of Machines by Humans
- Human-AI interaction (HAII) and Interactive Machine Learning (iML)
- HAII in Healthcare
- HAII in Cybersecurity
- Future Prospects

Modeling the human brain in the 1980s (black box)

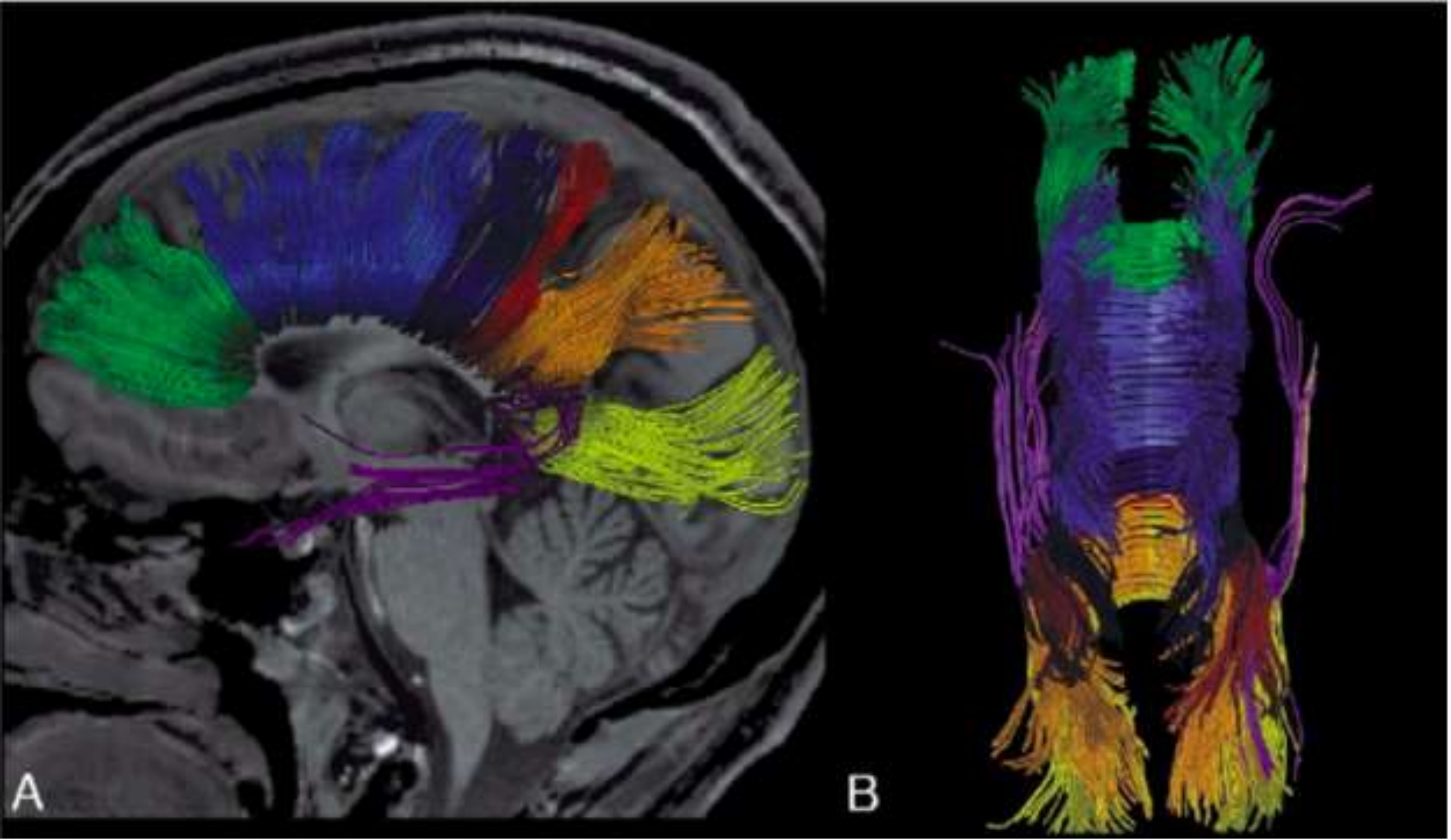
Timing is everything??

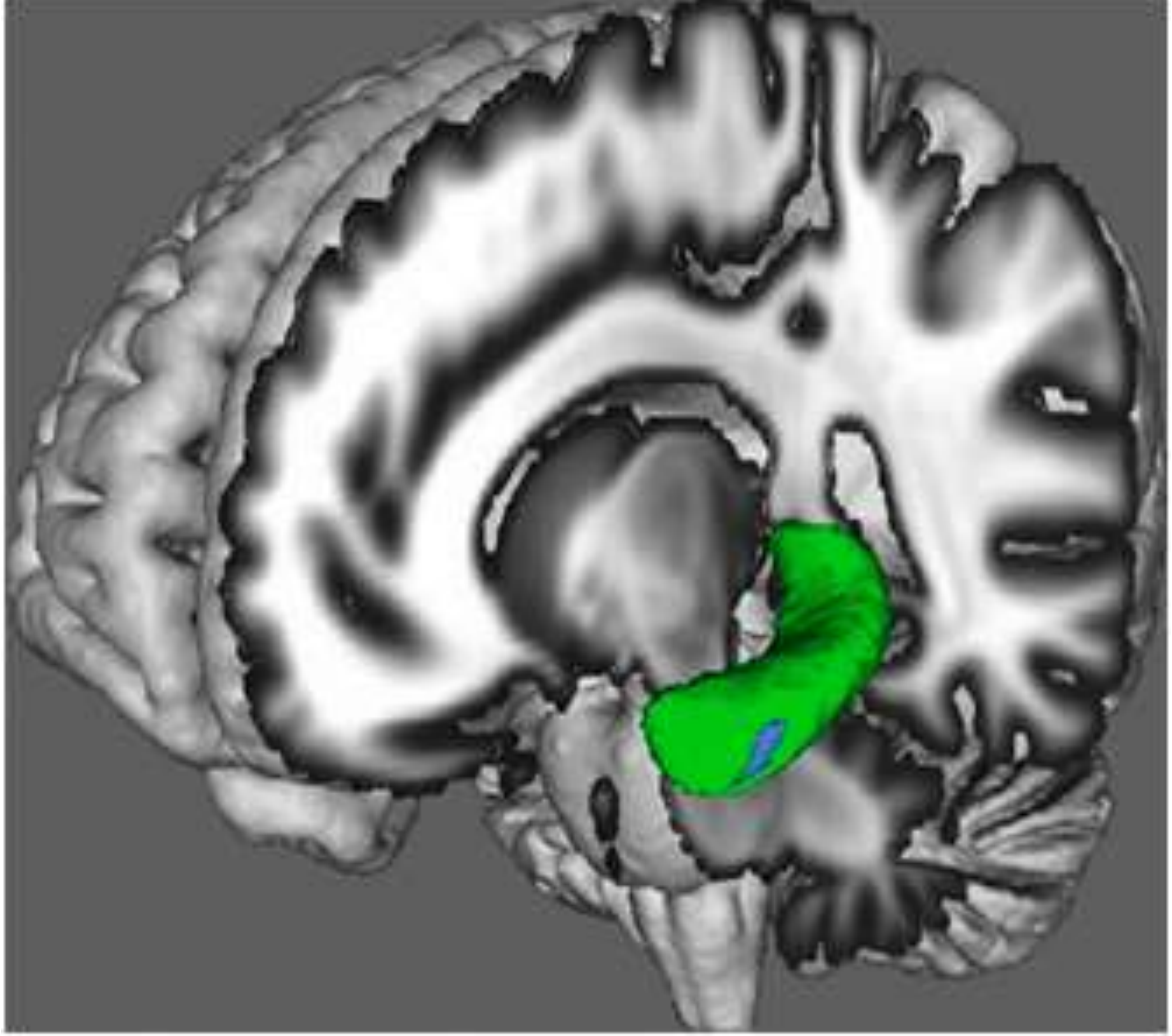

LONG-TERM MEMORY
$\delta_{LTM} = \infty$ sec, $\mu_{LTM} = \infty$, κ_{LTM} = Semantic
WORKING MEMORY
VISUAL IMAGE STORE
δ_{VIS} = 200 [70~1000] msec
μ_{VIS} = 17 [7~17] letters
κ_{VIS} = Iconic
AUDITORY IMAGE STORE
δ_{AIS} = 1500 [900~3500] msec
μ_{AIS} = 5 [4.4~6.2] letters
κ_{AIS} = Iconic
μ_{WM} = 3 [2.5~4.1] chunks
μ^{*}_{WM} = 7 [5~9] chunks
δ_{WM} = 7 [5~226] sec
δ_{WM} (1 chunk) = 73 [73~226] sec
δ_{WM} (3 chunks) = 7 [5~34] sec
κ_{WM} = acoustic
Perceptual Processor
τ_P = 100 [50~200] msec
Cognitive Processor
τ_C = 100 [25~170] msec
Motor Processor
τ_M - 100 [70~360] msec
Eye movement = 230 [70~700p msec

Our Wonderful Brain

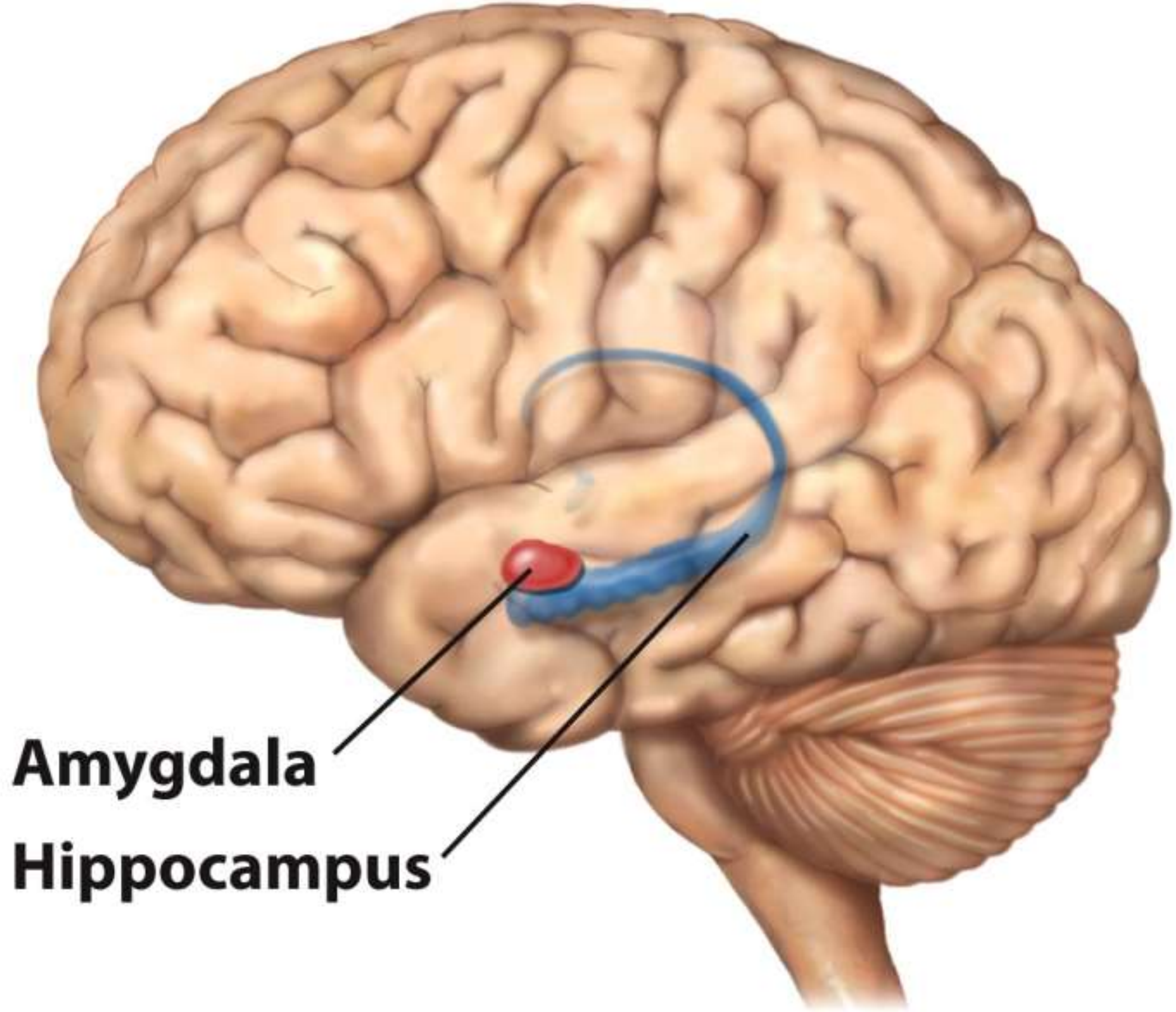


From the 1990s – Brain Scanning


A
B

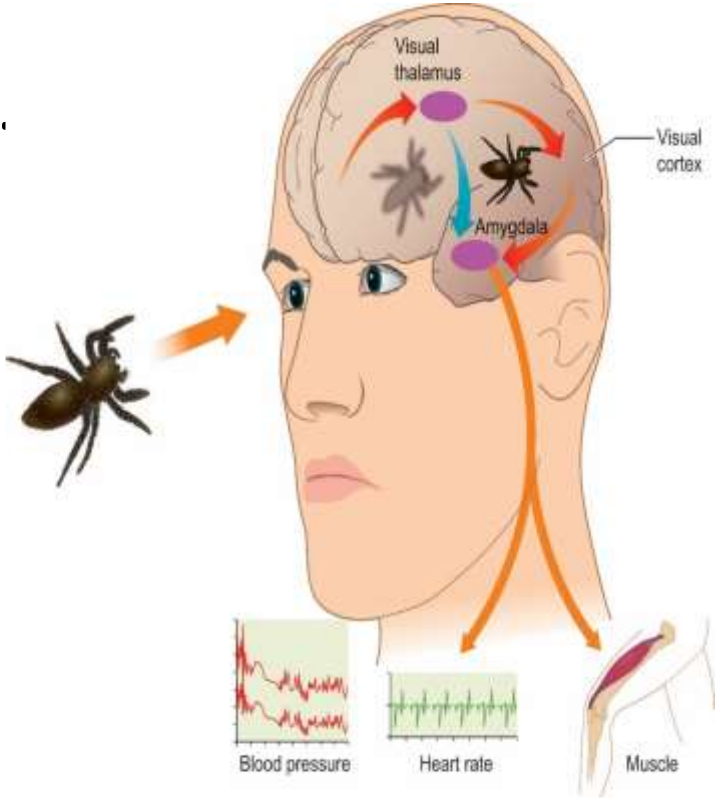


Location and circuitry of the amygdala



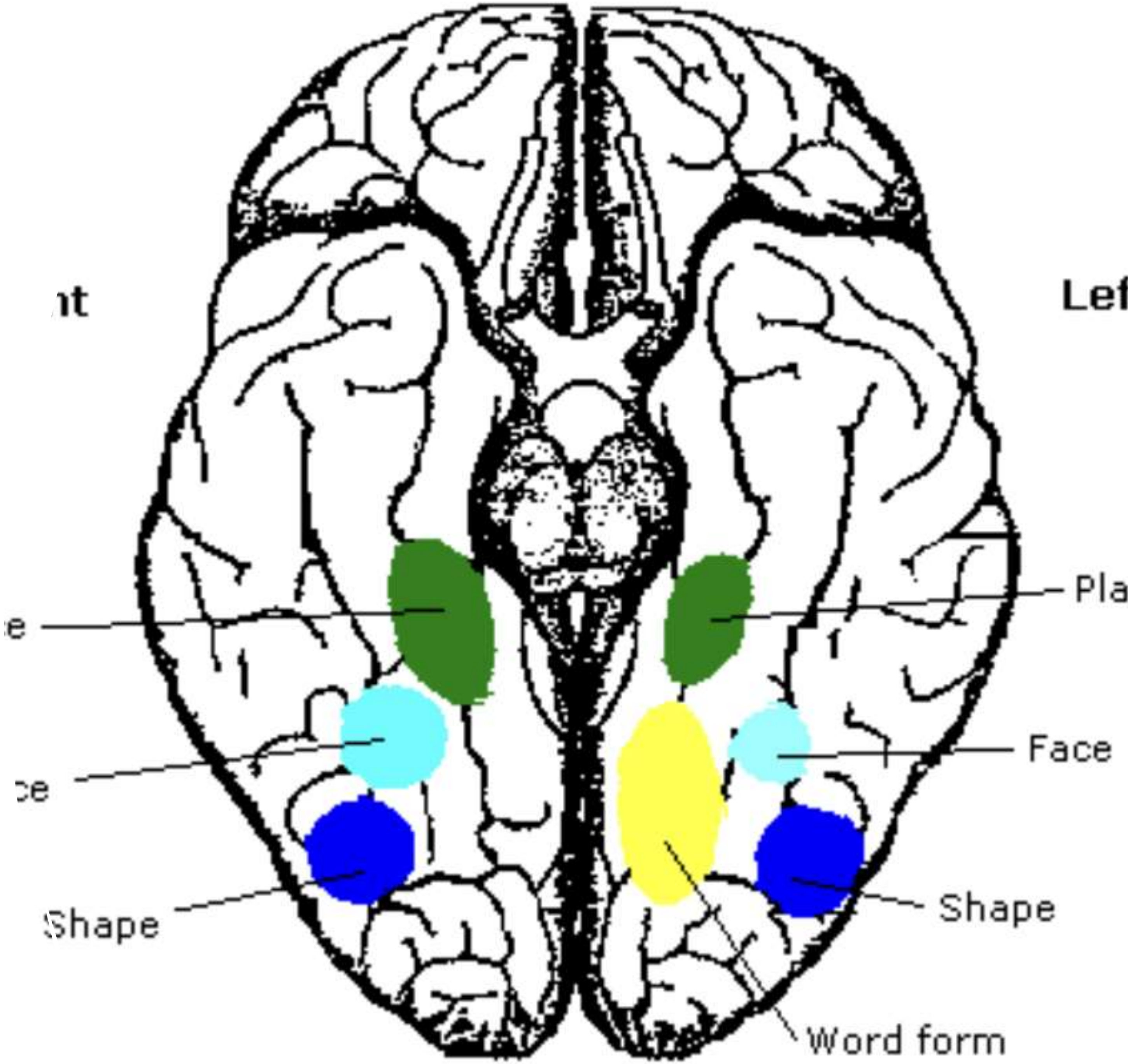
Connections to and from the Amygdala

- Sensory information is sent to the amygdala to enable emotional learning.
- Dual route model proposes two pathways:
  1. "Low road"
     - projects directly from the anterior thalamus to the amygdala.
     - acts as a "first alert" system, carrying a crude, preliminary sketch of basic properties of the stimulus.
  2. "High road"
     - connects the sensory areas of the cortex to the amygdala
     - provides a more comprehensive context for processing emotional information
     - gives rise to a slower affective reaction that takes into account the complexity and details of the situation



The Brain is a bunch of Co-Processors Connected by a Fibre Network

- Example Cortical Co-Processors
  - Frontal Eye Fields (in frontal cortex, plan eye movements)
  - Facial Fusiform Area (inferior temporal cortex, recognizes faces)
- Example Subcortical Co-Processors
  - Hippocampus (Memory)
  - Amygdala (Fear)



Building a Brain through Evolution

- Perception-Action Cycle
- Fast Survival - Fear, then other emotions
- Need to learn – Reward processing
- More complex learning – Memory (e.g., squirrel hiding nuts)
- More complex environments – Decision making
- Even more complex environmnets – AI co-processors and agents

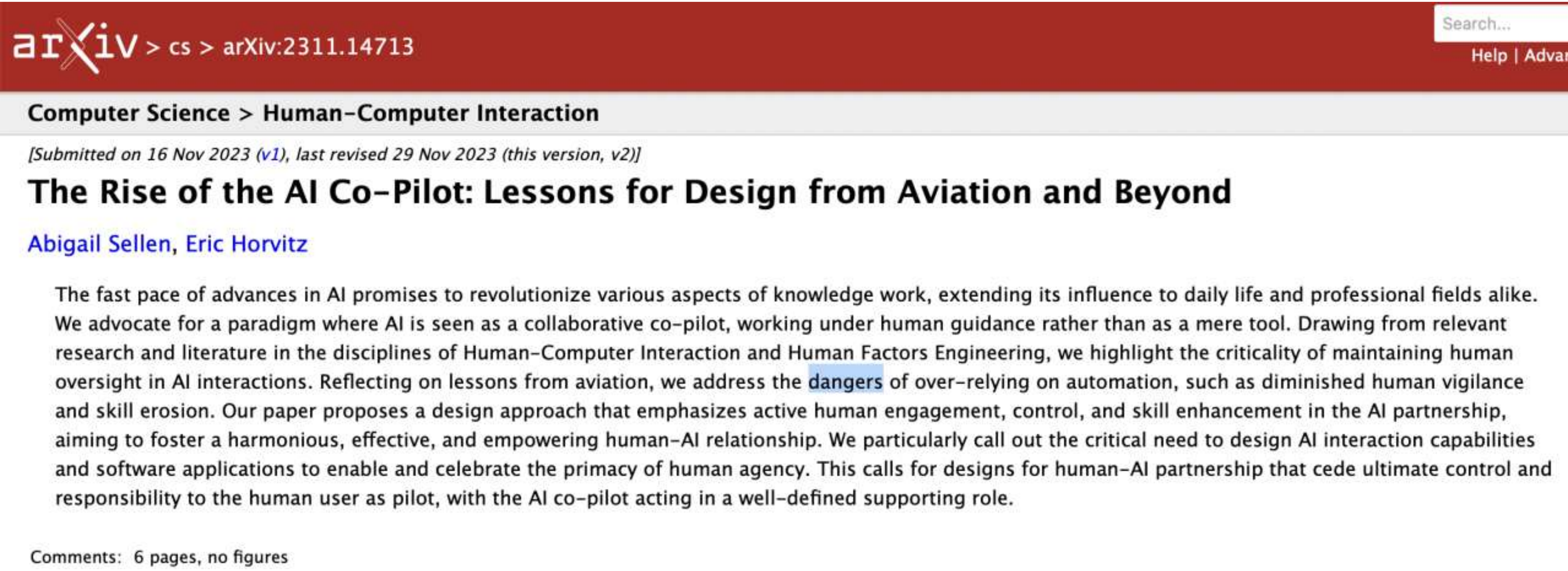
AI is another step in the evolution of the brain

- Just another co-processor
- A mobile phone is a hand-held co-processor
- A Voice agent is a wireless audio co-processor
- AI is a cognitive co-processor

Models of Human-AI Interaction

- AI is a Co-Pilot
- AI is an agent or “Master” (not over-lord)
- AI is a cognitive co-processor (ultimately an implant)

The Co-Pilot View



Computer Science > Human-Computer Interaction

[Submitted on 16 Nov 2023 (v1), last revised 29 Nov 2023 (this version, v2)]

The Rise of the AI Co-Pilot: Lessons for Design from Aviation and Beyond

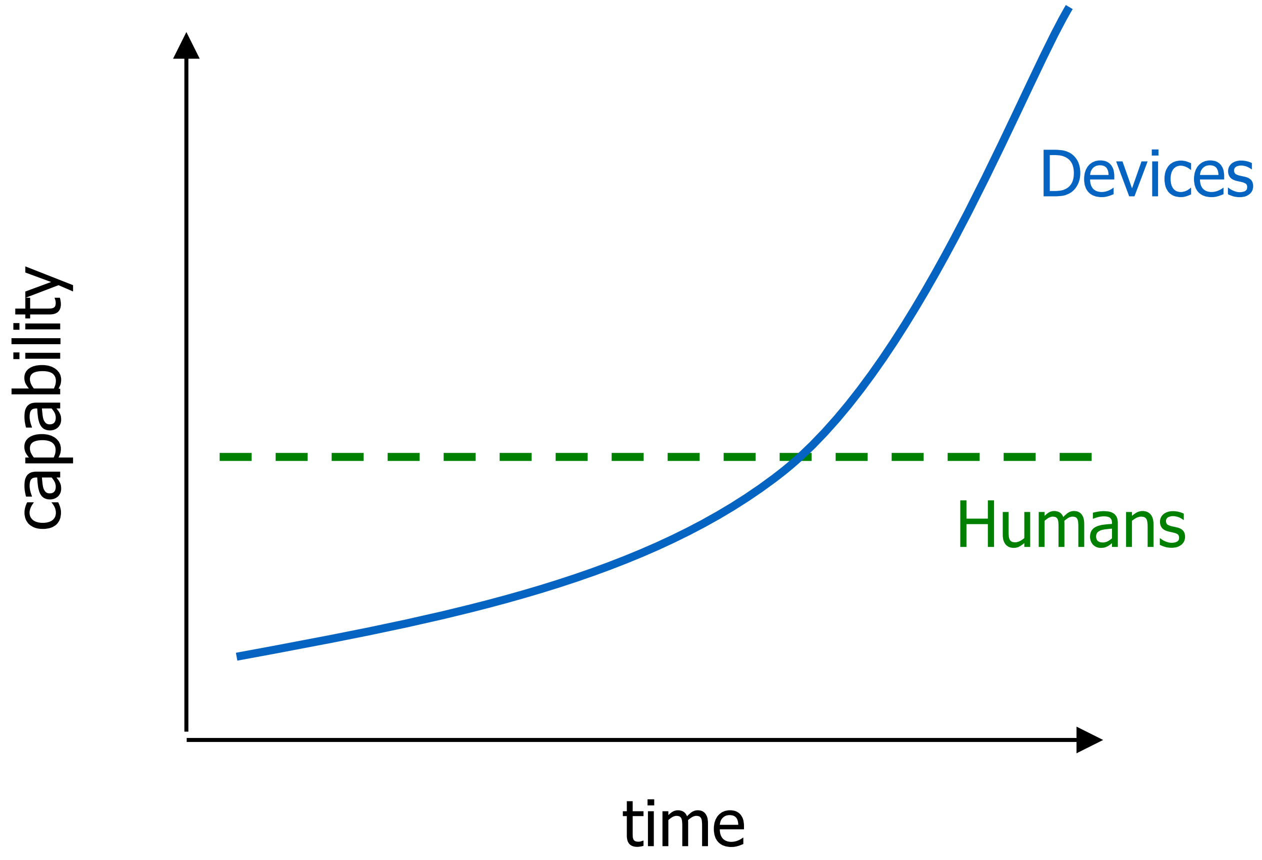
Abigail Sellen, Eric Horvitz

The fast pace of advances in AI promises to revolutionize various aspects of knowledge work, extending its influence to daily life and professional fields alike. We advocate for a paradigm where AI is seen as a collaborative co-pilot, working under human guidance rather than as a mere tool. Drawing from relevant research and literature in the disciplines of Human-Computer Interaction and Human Factors Engineering, we highlight the criticality of maintaining human oversight in AI interactions. Reflecting on lessons from aviation, we address the dangers of over-relying on automation, such as diminished human vigilance and skill erosion. Our paper proposes a design approach that emphasizes active human engagement, control, and skill enhancement in the AI partnership, aiming to foster a harmonious, effective, and empowering human-AI relationship. We particularly call out the critical need to design AI interaction capabilities and software applications to enable and celebrate the primacy of human agency. This calls for designs for human-AI partnership that cede ultimate control and responsibility to the human user as pilot, with the AI co-pilot acting in a well-defined supporting role.

Comments: 6 pages, no figures

Model Mastery and Servitude: The Surprising Resilience of Apprentice Models and Outperformed Experts

Human Capability



Buxton, W. (2001). Less is More (More or Less), in P. Denning (Ed.), The Invisible Future: The seamless integration of technology in everyday life. New York: McGraw Hill, 145 – 179.

Clinical versus actuarial judgment

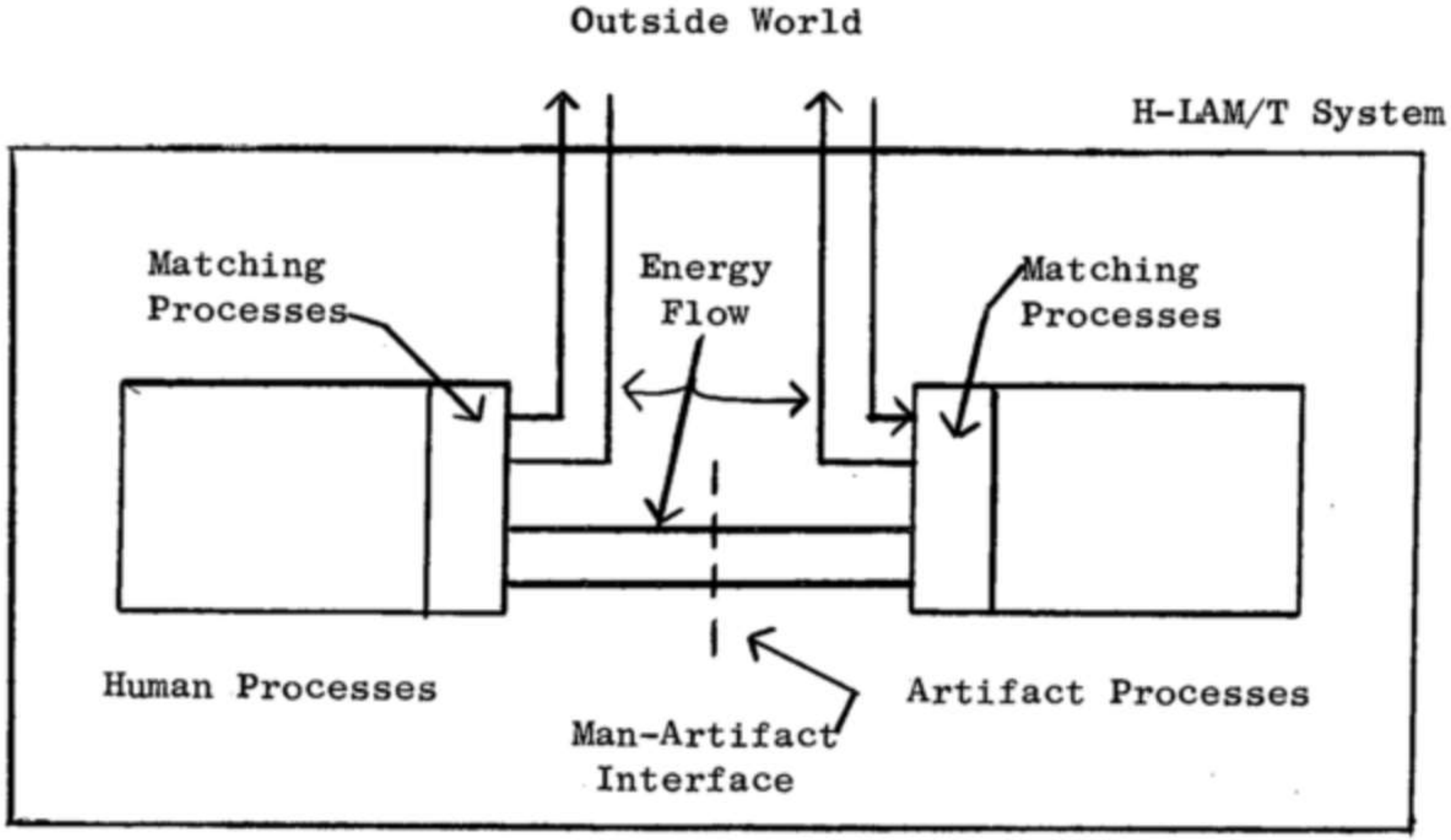
- Paul Meehl (1954) first addressed the question: Which is better?

"…it is clear that the dogmatic, complacent assertion sometimes heard from clinicians that 'naturally' clinical prediction, being based on 'real understanding' is superior, is simply not justified by the facts to date".

Visions of Human-AI interaction in the 1960s

Augmenting Human Intellect:
A Conceptual Framework

By Douglas C. Engelbart
October 1962

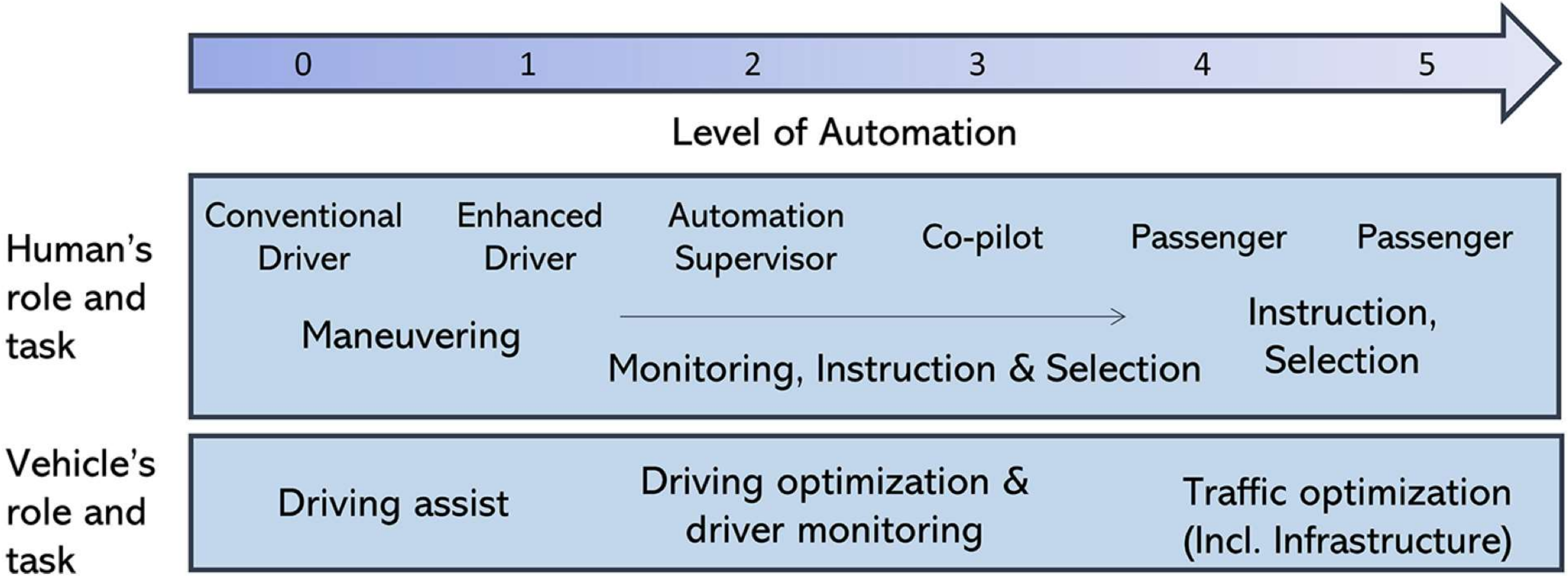


Machine Augmentation and Human Augmentation

- Automated Driving is an example where the human augments machine capability
- But the vision was different in the 1960s
  - *Automation would reduce work requirements and give us too much leisure time*
  - *AI and other technologies would make us more powerful by augmenting us*
- Didn’t happen. We seem to be living in an age where the tools we have developed are concentrating power in the hands of fewer and fewer people, leaving many people feeling disempowered rather than augmented
- So what can we learn from automated driving?

The Changing Human Role In Automated Vehicles

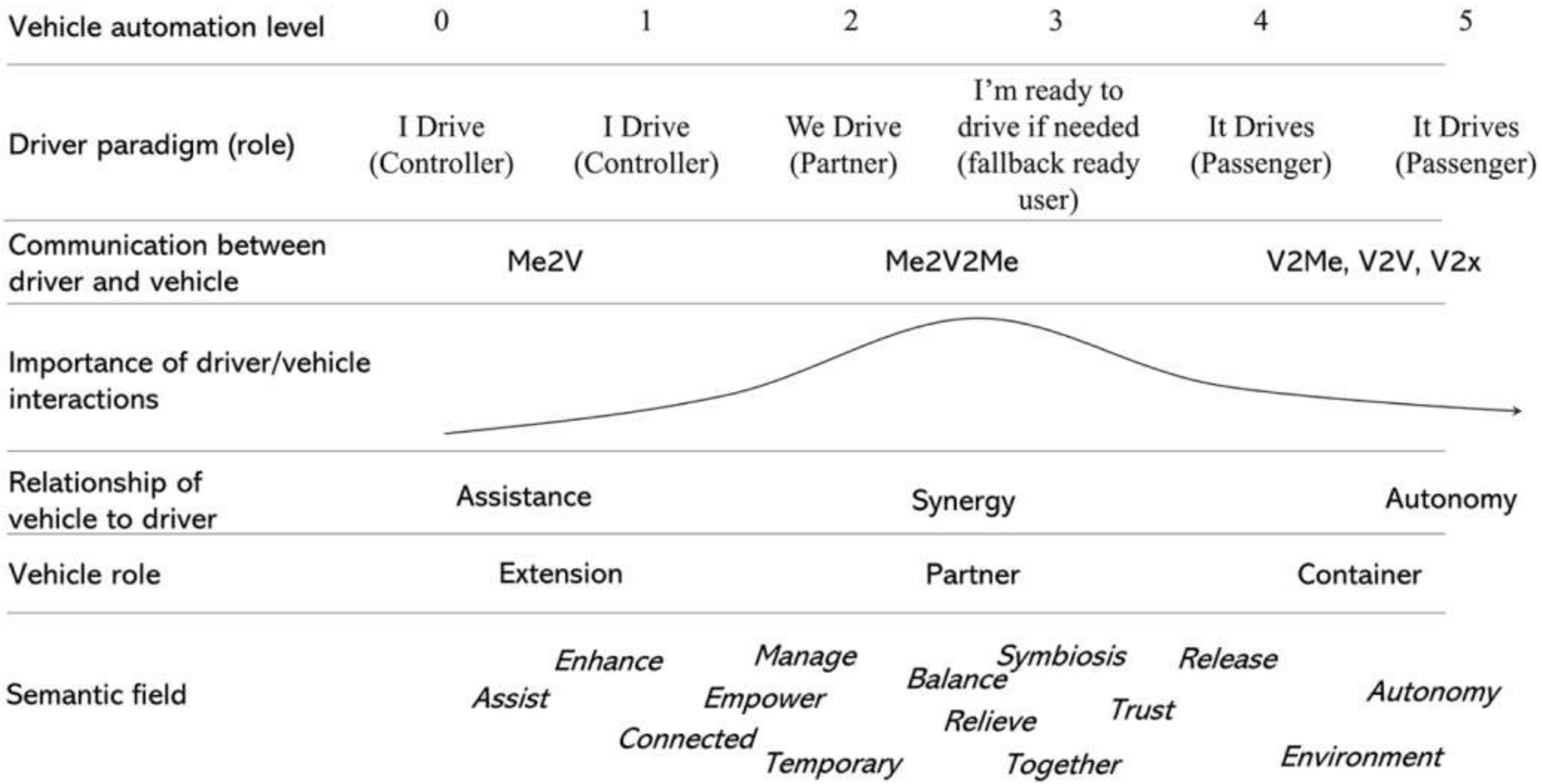
In the 1980s, the Personal Computer made the user pre-eminent,
in the future Human-AI interaction will be increasingly important



Level of Automation

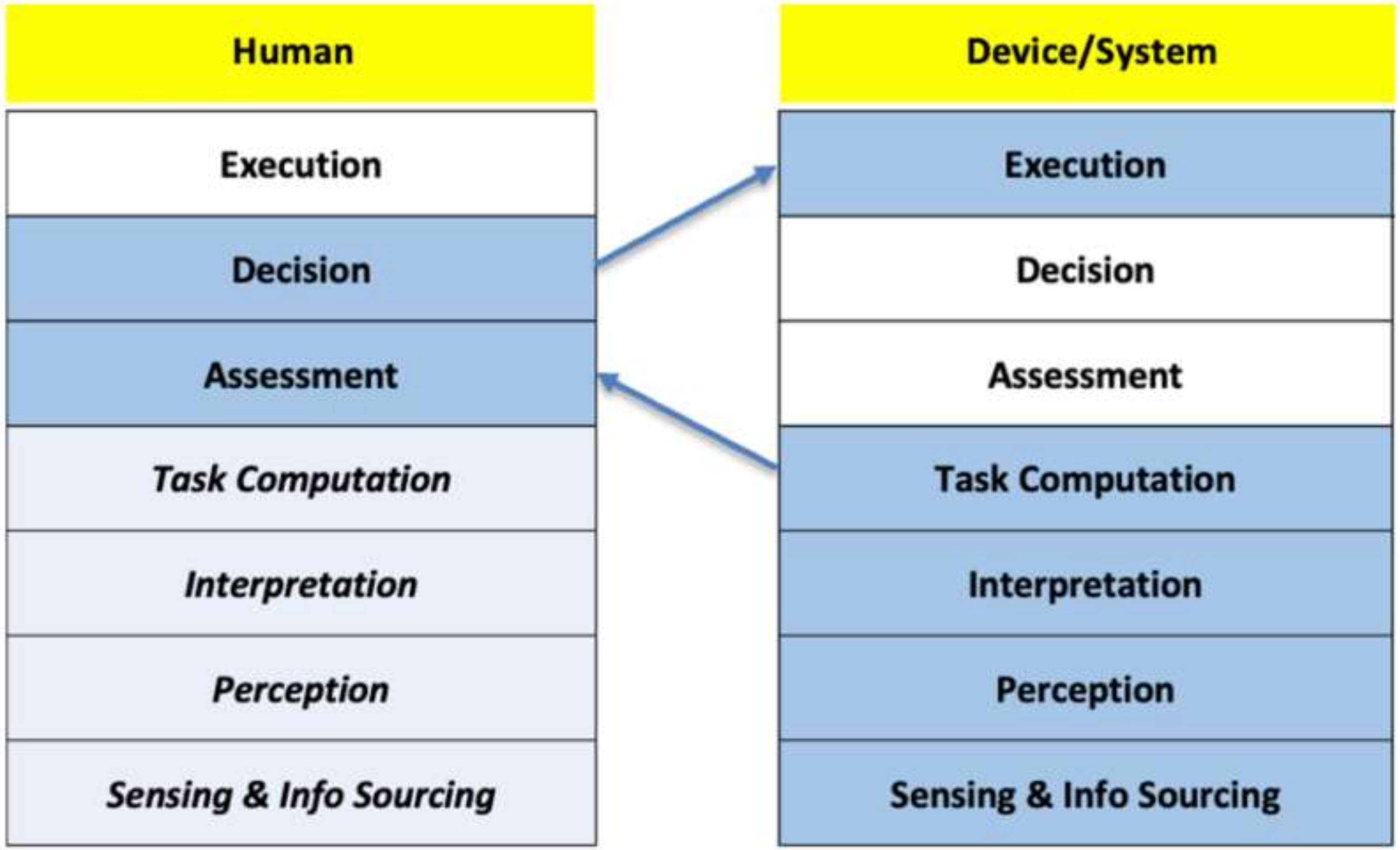
	0	1	2	3	4	5
Human's role and task	Conventional Driver	Enhanced Driver	Automation Supervisor	Co-pilot	Passenger	Passenger
	Maneuvering		Monitoring, Instruction & Selection			Instruction, Selection
Vehicle's role and task	Driving assist		Driving optimization & driver monitoring		Traffic optimization (Incl. Infrastructure)	

When Human-AI Interaction Matters



Vehicle automation level	0	1	2	3	4	5
Driver paradigm (role)	I Drive (Controller)	I Drive (Controller)	We Drive (Partner)	I'm ready to drive if needed (fallback ready user)	It Drives (Passenger)	It Drives (Passenger)
Communication between driver and vehicle	Me2V		Me2V2Me		V2Me, V2V, V2x	
Importance of driver/vehicle interactions						
Relationship of vehicle to driver	Assistance		Synergy			Autonomy
Vehicle role	Extension		Partner		Container	
Semantic field	Assist	Enhance Connected	Manage Empower Temporary	Balance Symbiosis Relieve Trust Together	Release	Autonomy Environment

Cooperative Task Performance with AI Agent

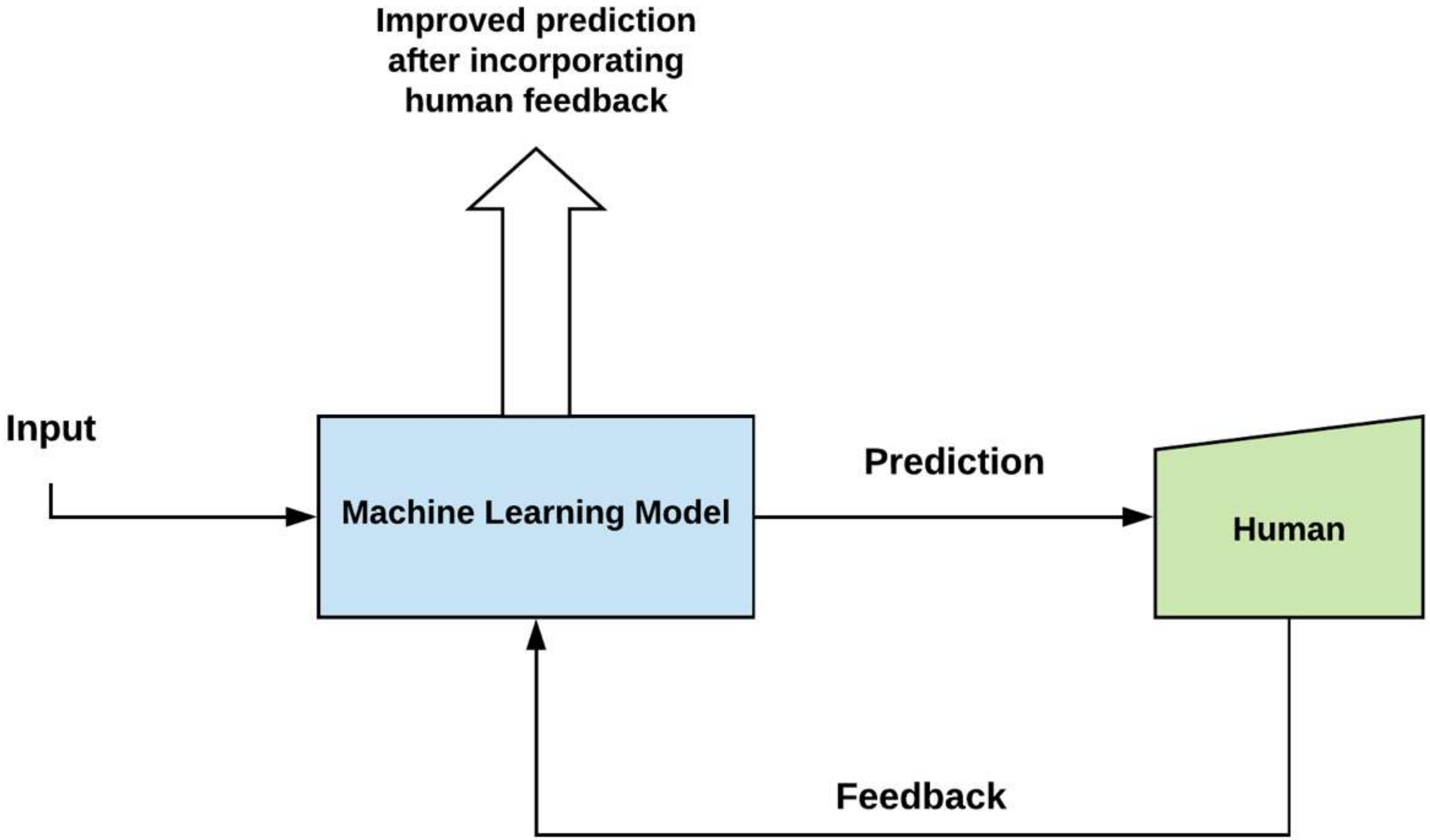


Human	Device/System
Execution	Execution
Decision	Decision
Assessment	Assessment
Task Computation	Task Computation
Interpretation	Interpretation
Perception	Perception
Sensing & Info Sourcing	Sensing & Info Sourcing

Interactive Machine Learning (iML)

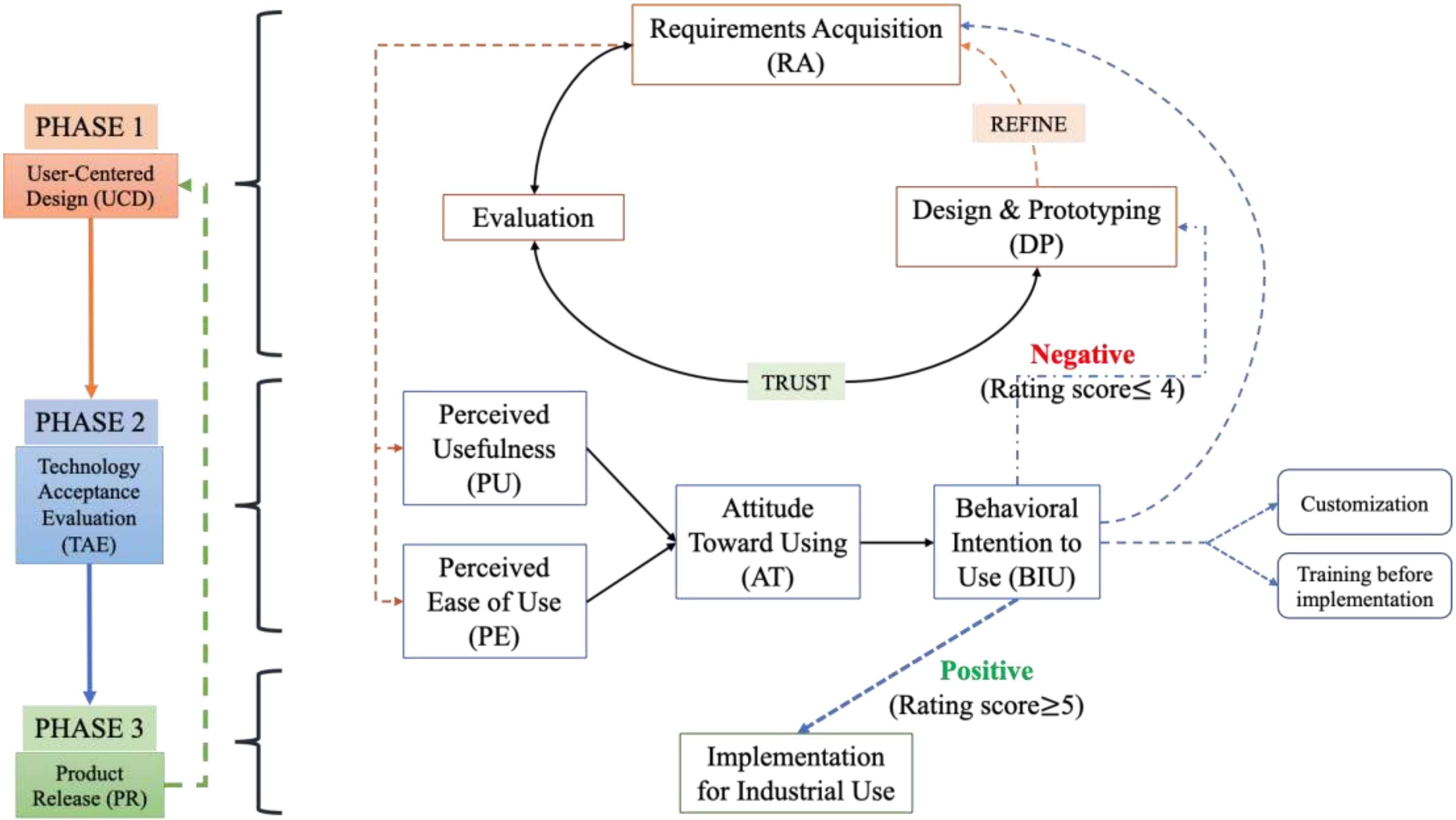
Training an AI system

A Simple View of Machine Learning

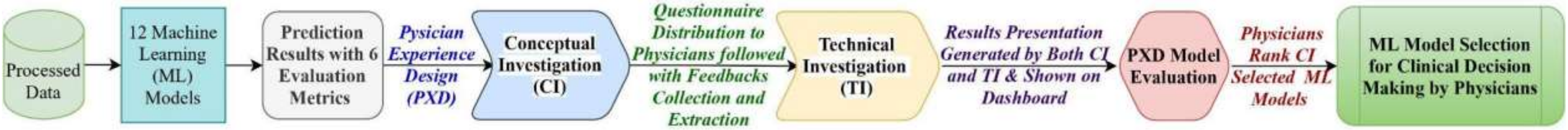


iML Case Study in Healthcare

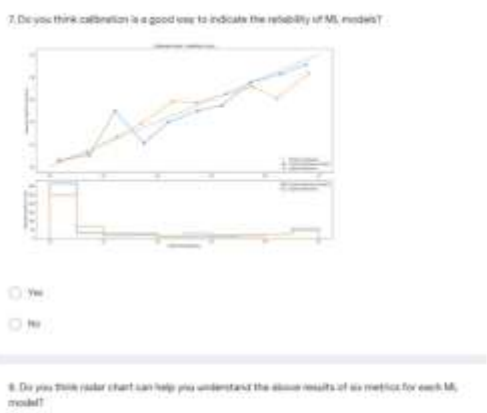
Overall HCAI Framework


PHASE 1
User-Centered Design (UCD)
PHASE 2
Technology Acceptance Evaluation (TAE)
PHASE 3
Product Release (PR)
Requirements Acquisition (RA)
Evaluation
Design & Prototyping (DP)
REFINE
TRUST
Negative
(Rating score≤ 4)
Perceived Usefulness (PU)
Perceived Ease of Use (PE)
Attitude Toward Using (AT)
Behavioral Intention to Use (BIU)
Customization
Training before implementation
Positive
(Rating score≥5)
Implementation for Industrial Use

Physician Experience Design (PXD)


Processed Data
12 Machine Learning (ML) Models
Prediction Results with 6 Evaluation Metrics
Pysician Experience Design (PXD)
Conceptual Investigation (CI)
Questionnaire Distribution to Physicians followed with Feedbacks Collection and Extraction
Technical Investigation (TI)
Results Presentation Generated by Both CI and TI & Shown on Dashboard
PXD Model Evaluation
Physicians Rank CI Selected ML Models
ML Model Selection for Clinical Decision Making by Physicians

Making machine learning models more usable with physician experience design



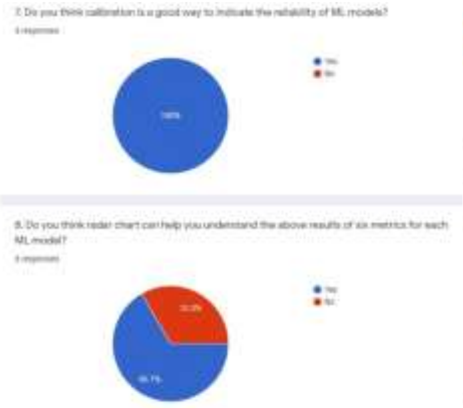
7. Do you think calibration is a good way to indicate the reliability of ML models?

○ Yes

○ No

8. Do you think radar chart can help you understand the above results of six metrics for each ML model?

(a) Partial questionnaire.



7. Do you think calibration is a good way to indicate the reliability of ML models?

8. Do you think radar chart can help you understand the above results of six metrics for each ML model?

(b) Partial answers.

Physician Experience Design

(a) PXD dashboard with three levels of dropdown menus.

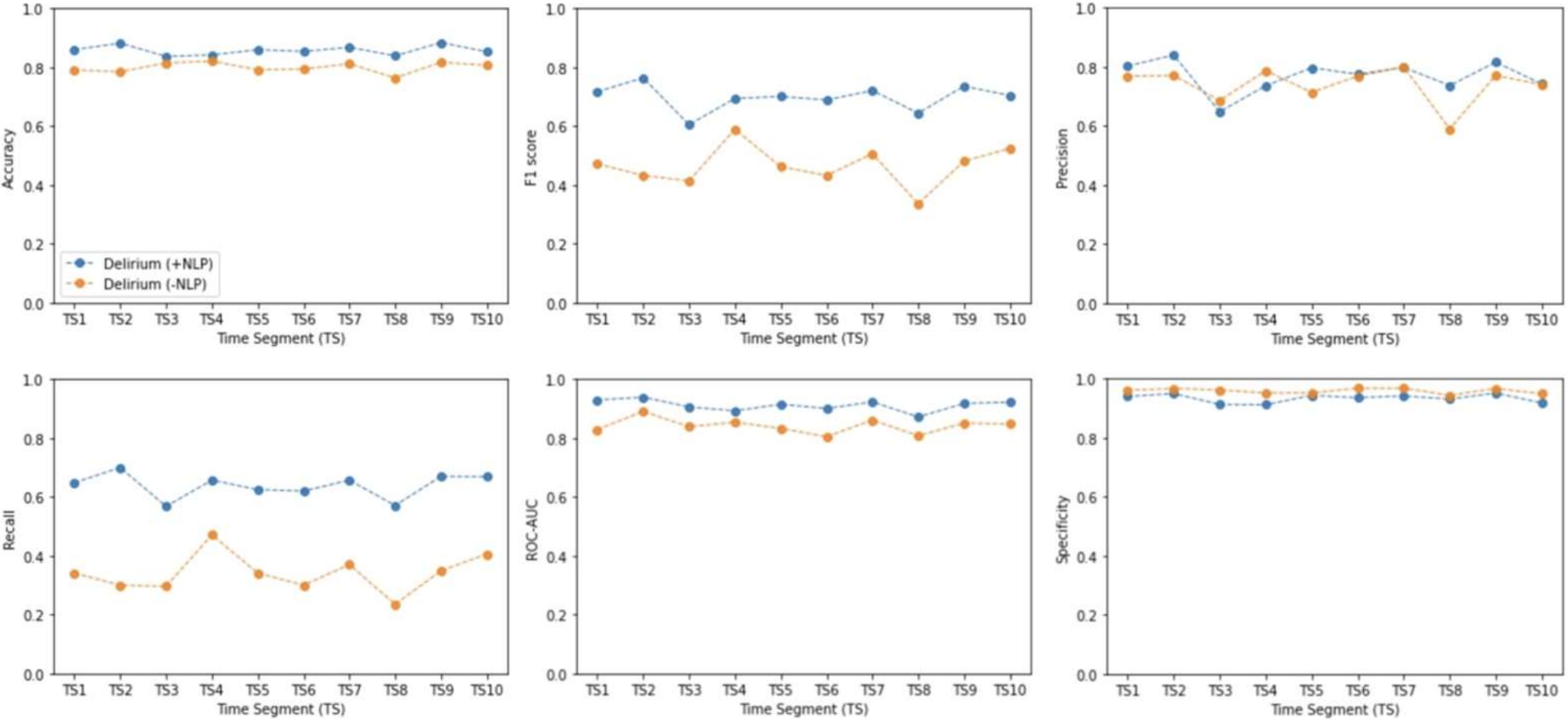
(b) Two "readme" buttons for users quickly start.

Sample Figures and Dashboards

Example: Using Delirium Sentiment Analysis to Improve ML prediction

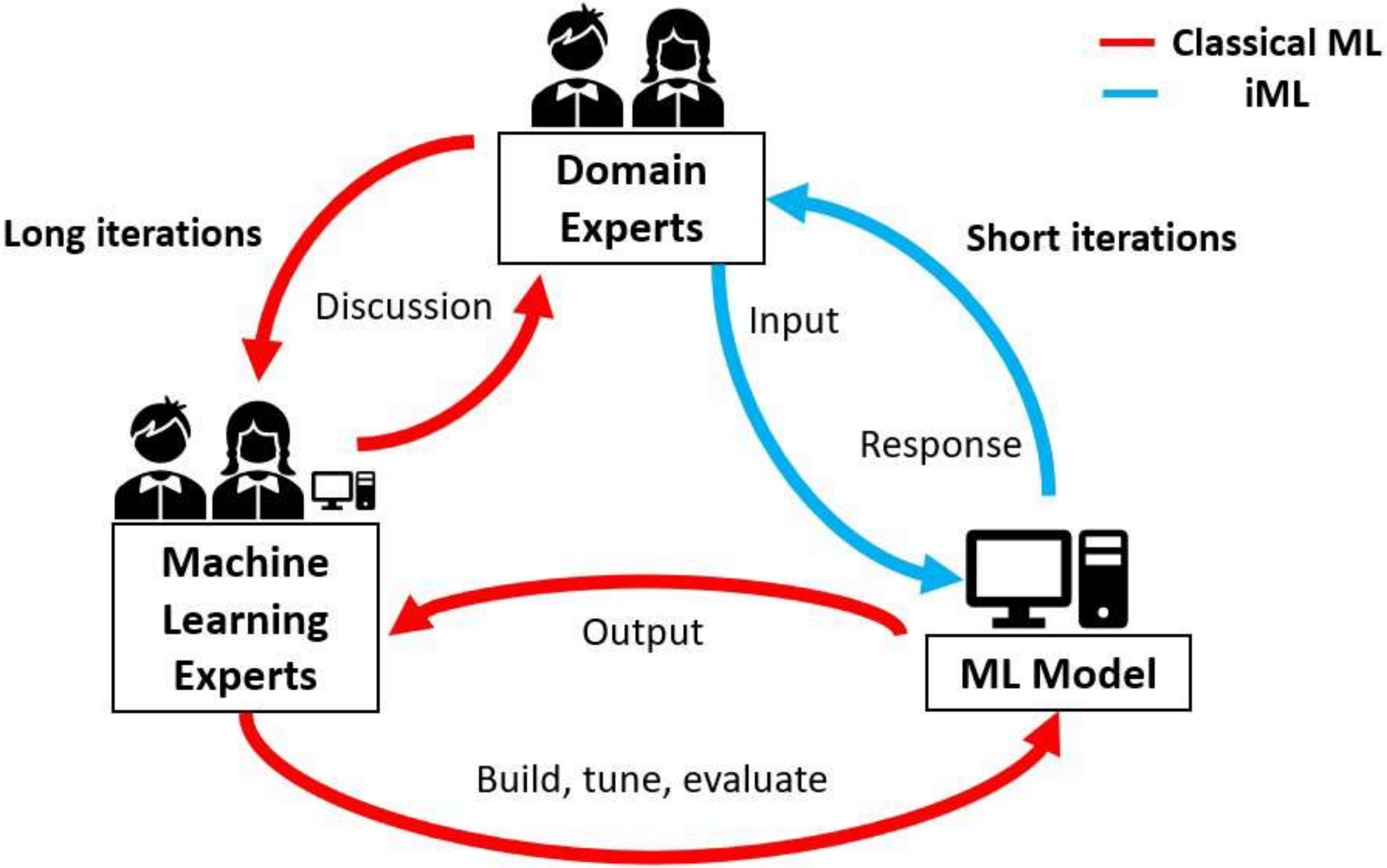
- Look for the words that clinicians use to describe cases that are later labeled as having delirium
- Use ML training to create a feature that measures “delirium sentiment”
- Measure how much delirium sentiment boosts ML prediction of delirium

ML performance (with and without Sentiment Feature) over Ten Time Segments

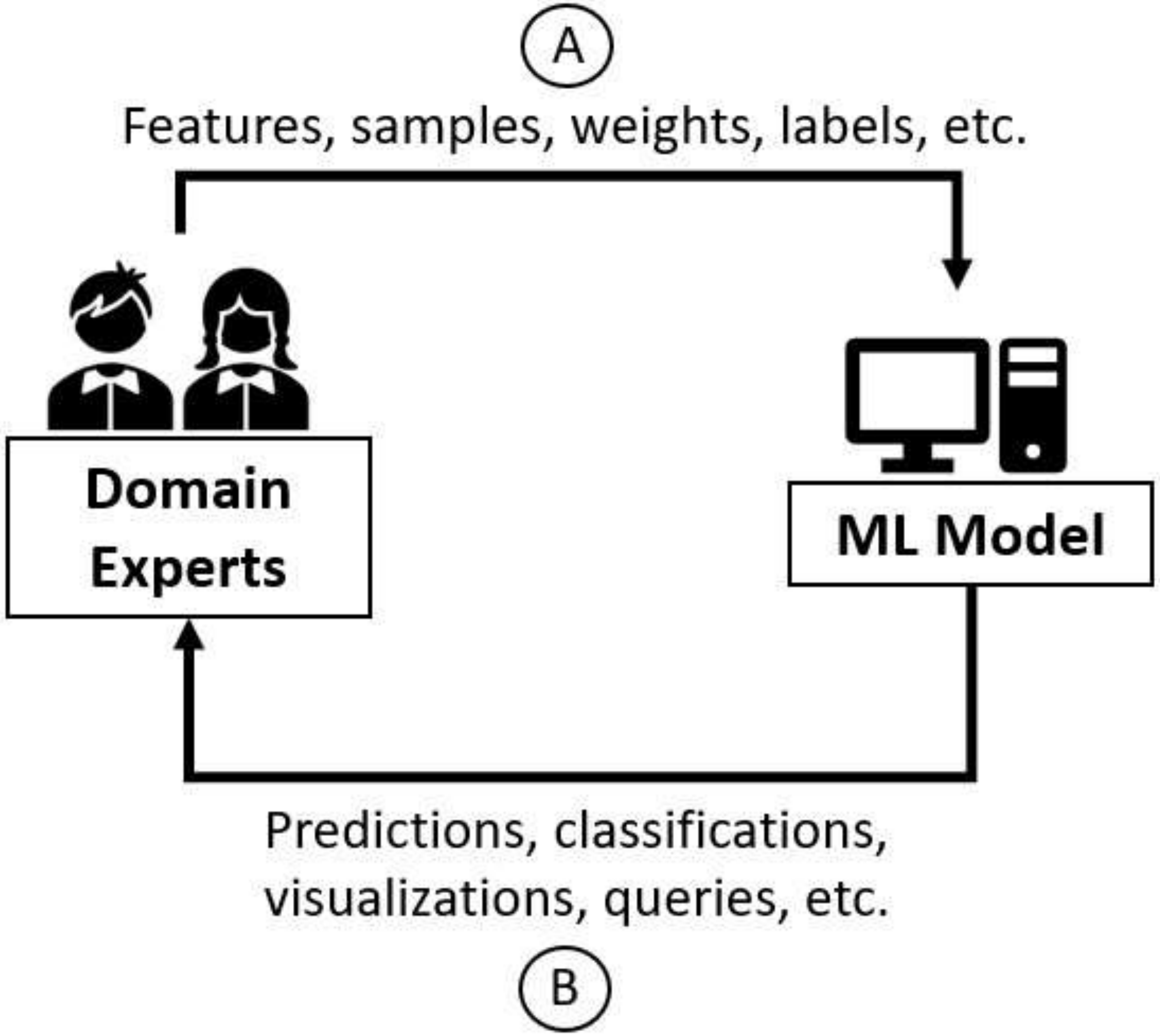


iML in Cybersecurity:
A Data Exfiltration Case Study

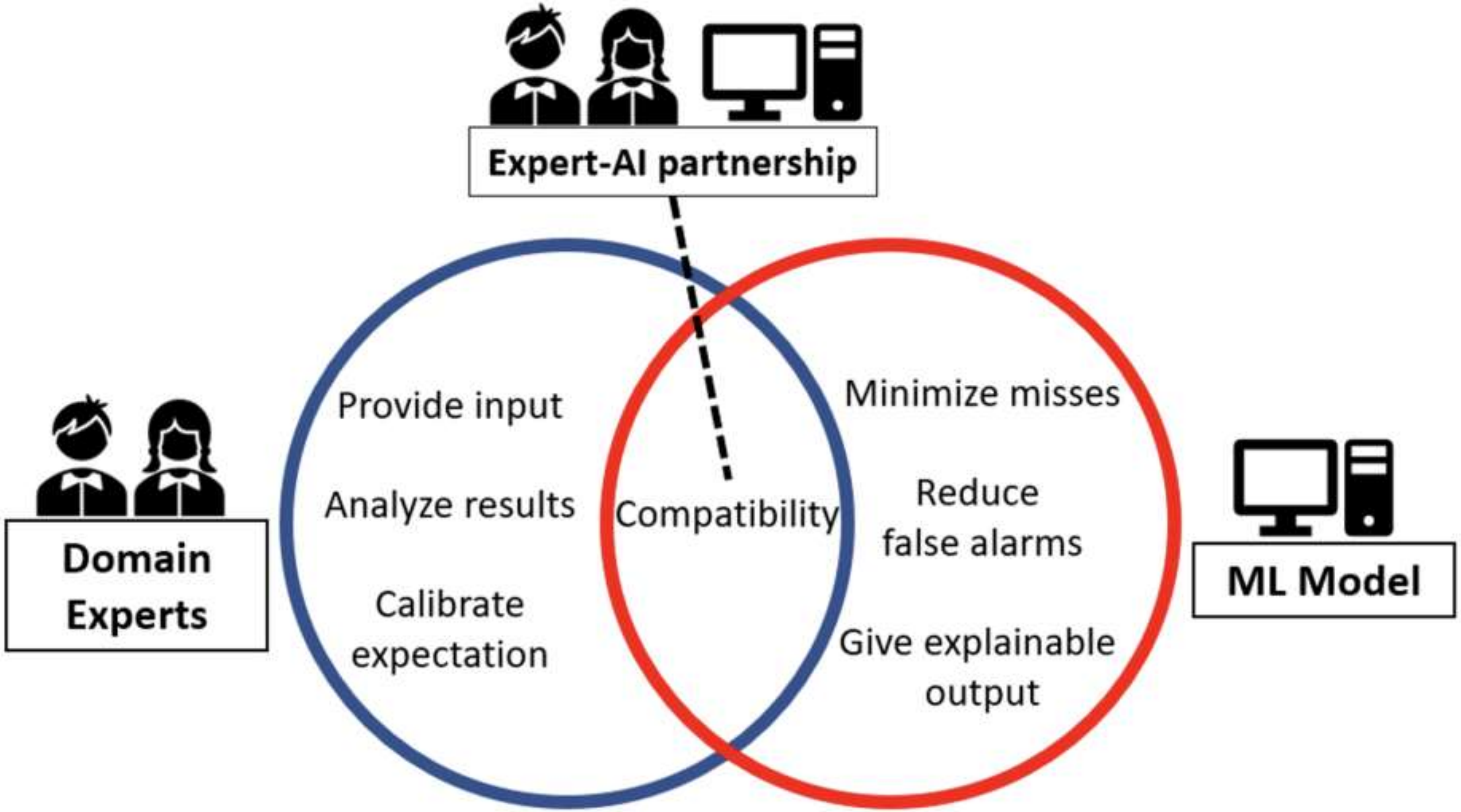
Interactive ML (iML) vs. Traditional ML



Tasks and Feedback for Domain Experts


A
Features, samples, weights, labels, etc.
Domain Experts
ML Model
Predictions, classifications, visualizations, queries, etc.
B

Data Exfiltration (Cybersecurity)

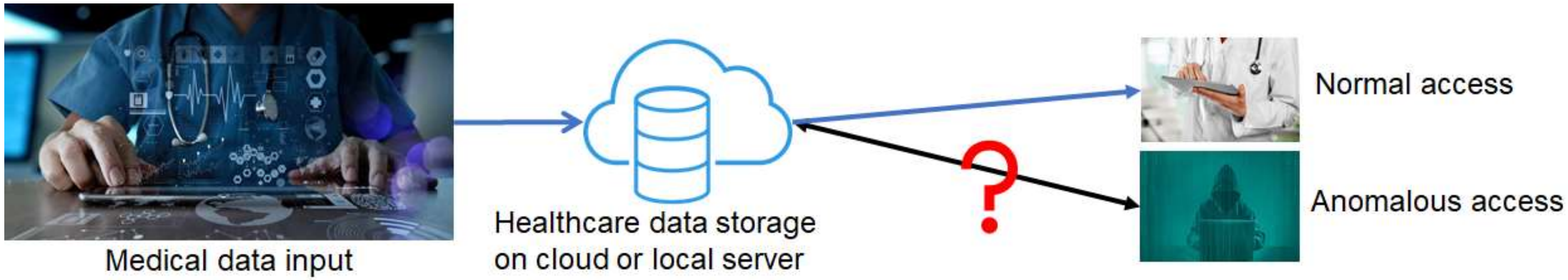


Use Cases

- A compromised account may have the following anomalous behaviours:
    - Unusual login IPs (i.e., login from multiple continent/countries within a day)
    - Suddenly starting to send emails to too many other employees
    - Unusual login time
    - Subjects concerning topics unrelated to their business functions
    - Sharing files with external accounts frequently

A Dilemma Handling Data Exfiltration Threats

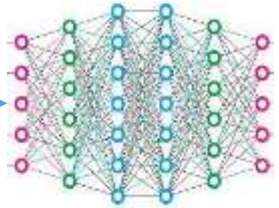
- Anomalous access can be performed by:
  - Abnormal benign system users
  - Malicious system users (can be insiders)
- Distinguish between abnormal benign and malicious users is troublesome because:
  - It **takes too long** for cybersecurity experts to investigate **manually**.
  - It is **hard for the machine** to identify maliciousness in anomalous activities automatically.



Detecting malicious user activities using interactive ML (iML)

System user activities

Machine learning algorithm to detect anomalous activities

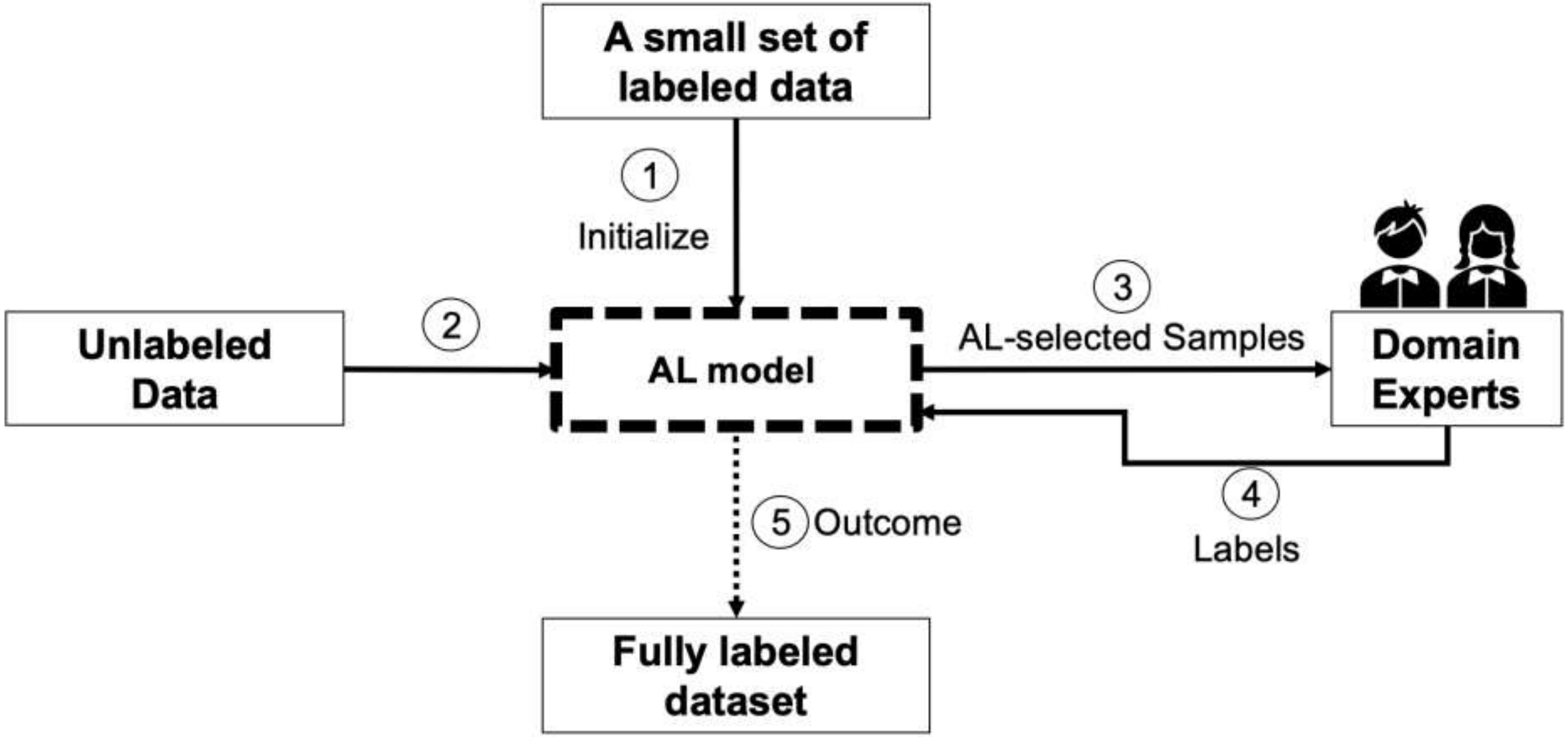


+

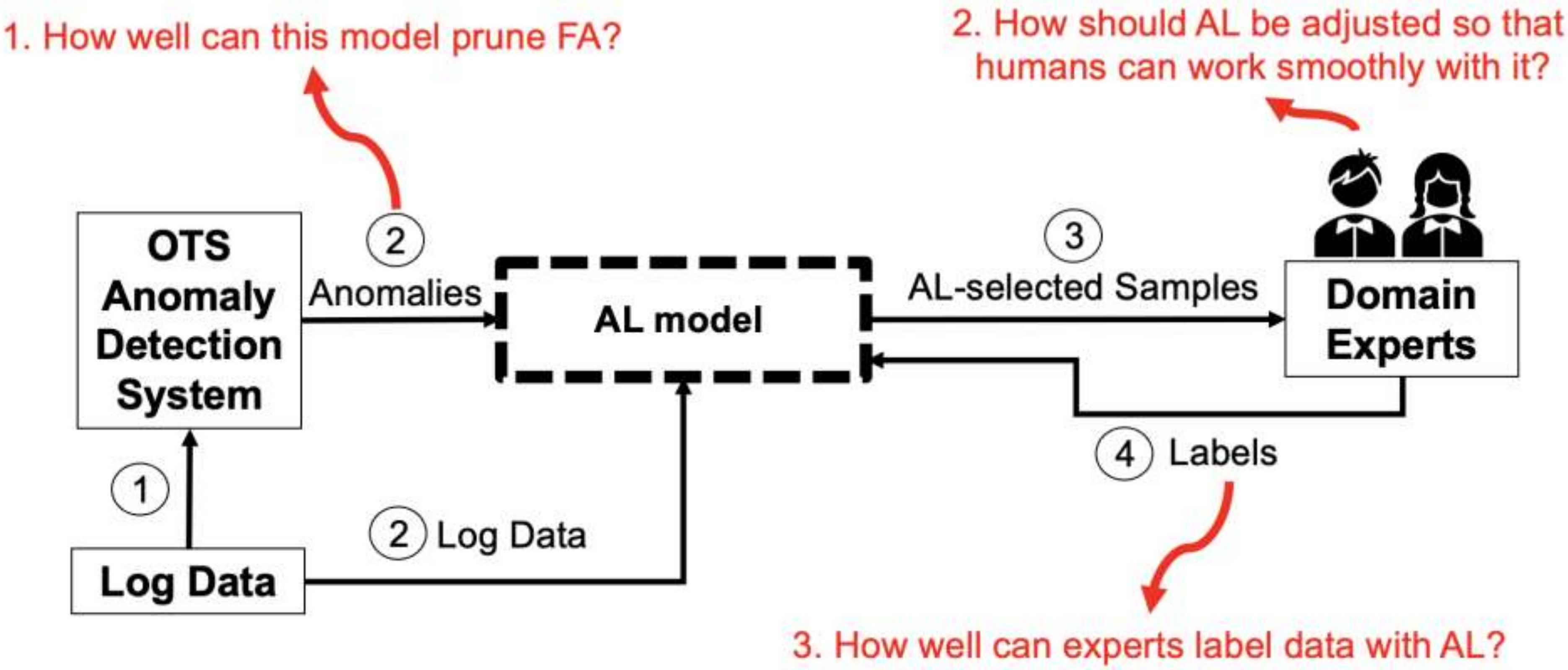
Domain experts using visualization tool to identify maliciousness

A model that can distinguish between normal users and malicious users

Active Learning (AL) with Experts

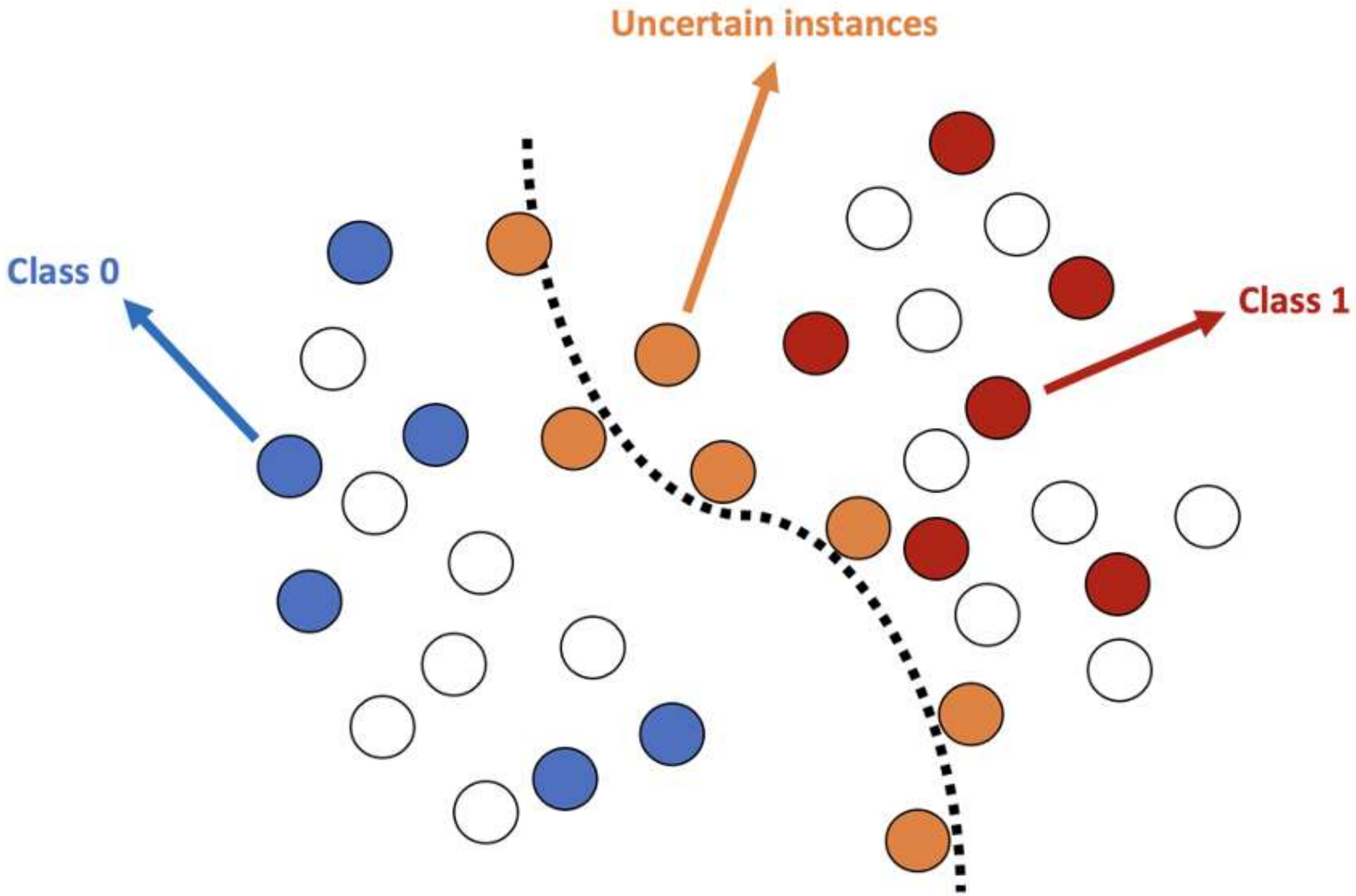


Research Questions

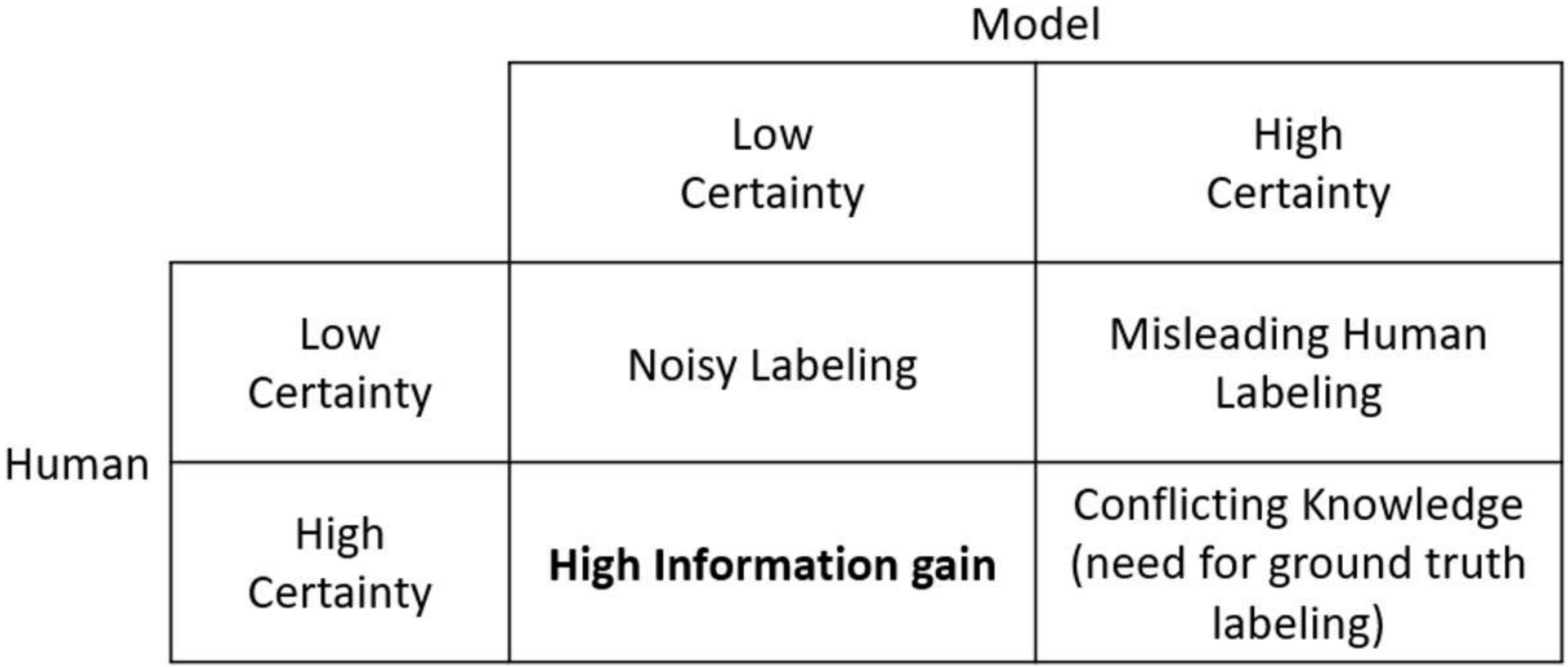


1. How well can this model prune FA?
2. How should AL be adjusted so that humans can work smoothly with it?
OTS Anomaly Detection System
2 Anomalies
AL model
3 AL-selected Samples
Domain Experts
1
Log Data
2 Log Data
4 Labels
3. How well can experts label data with AL?

AL in a Binary Classification Task



Focusing on High Information Gain in AL



		Model	
		Low Certainty	High Certainty
Human	Low Certainty	Noisy Labeling	Misleading Human Labeling
	High Certainty	**High Information gain**	Conflicting Knowledge (need for ground truth labeling)

Detection of Anomalies in Finance Services Company

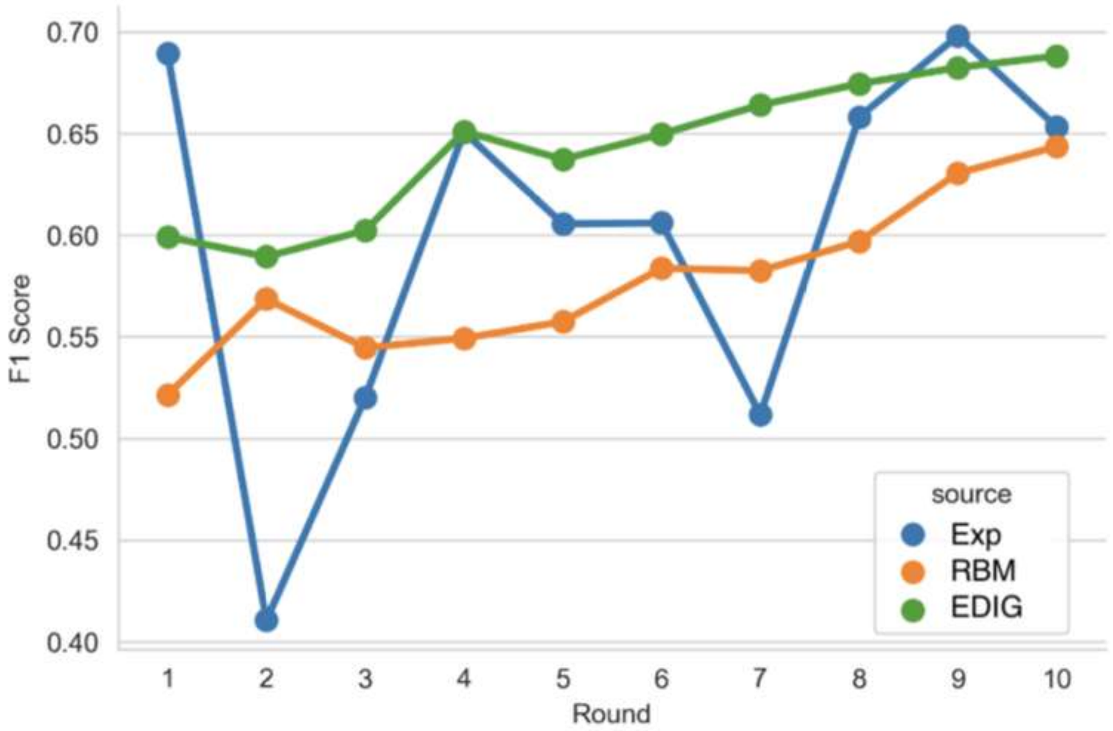


Figure 2: Average expert performance in F1 score compared with the two models they trained in each round

Overall learning (beginning vs. end)

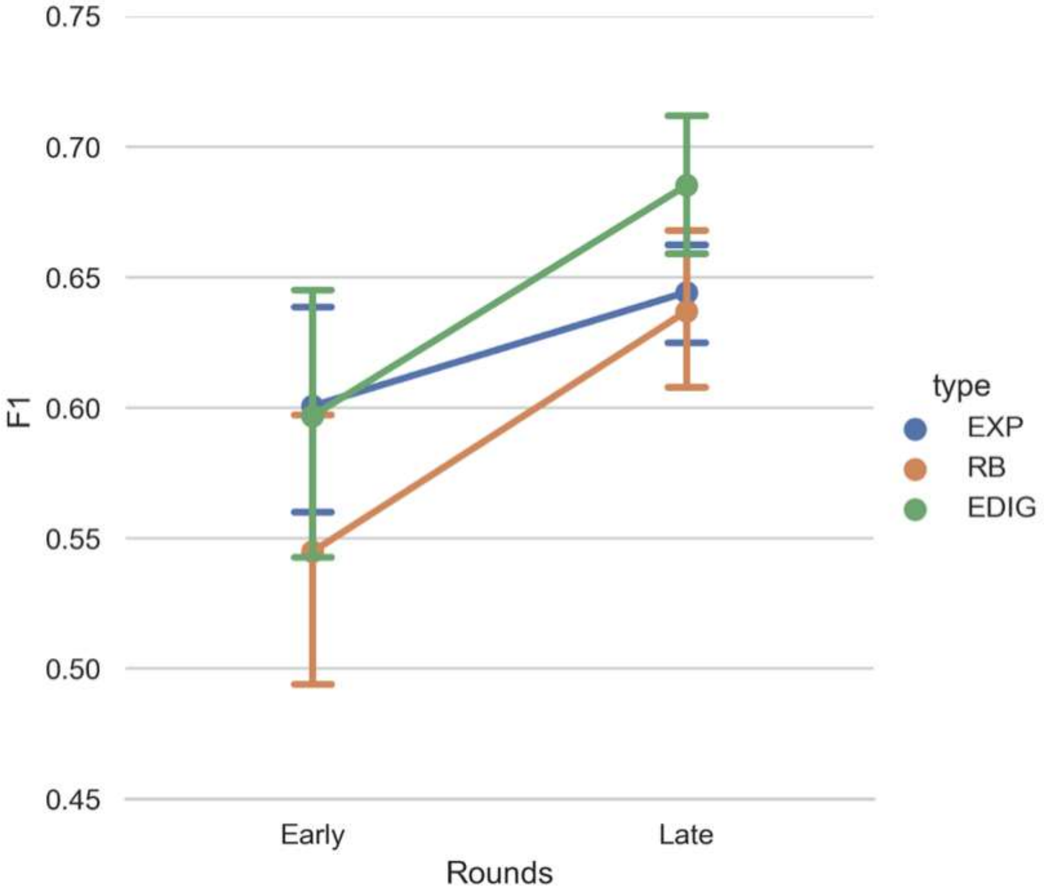


Figure 3: A Comparison for Expert and Model Performance in the first three ("Early") and last three ("Late") rounds of training (error bars show one standard error)

17
18
ORDERS
Caution
Emergency Department

Key Human Factors Issues

- Trust and Reliance
- Situation Awareness
- Workload
- Supervisory Control
- Safety
- Interruptability and Distraction

Conclusions

- Human-AI interaction is one of the central problems of our time
- For the moment, the focus is on augmenting AI, not on augmenting humans
- Why focus on iML?
    - ML is a set of powerful tools with good results across many applications
    - PXD is critical to effective use of ML in healthcare
    - Enhanced AL with good anomaly labeling UX is a promising approach for iML in Cybersecurity
    - There are undoubtedly many other applications where iML will be useful