Running Up That HILL

How Human-In-The-Loop Learning Leads to Better AI

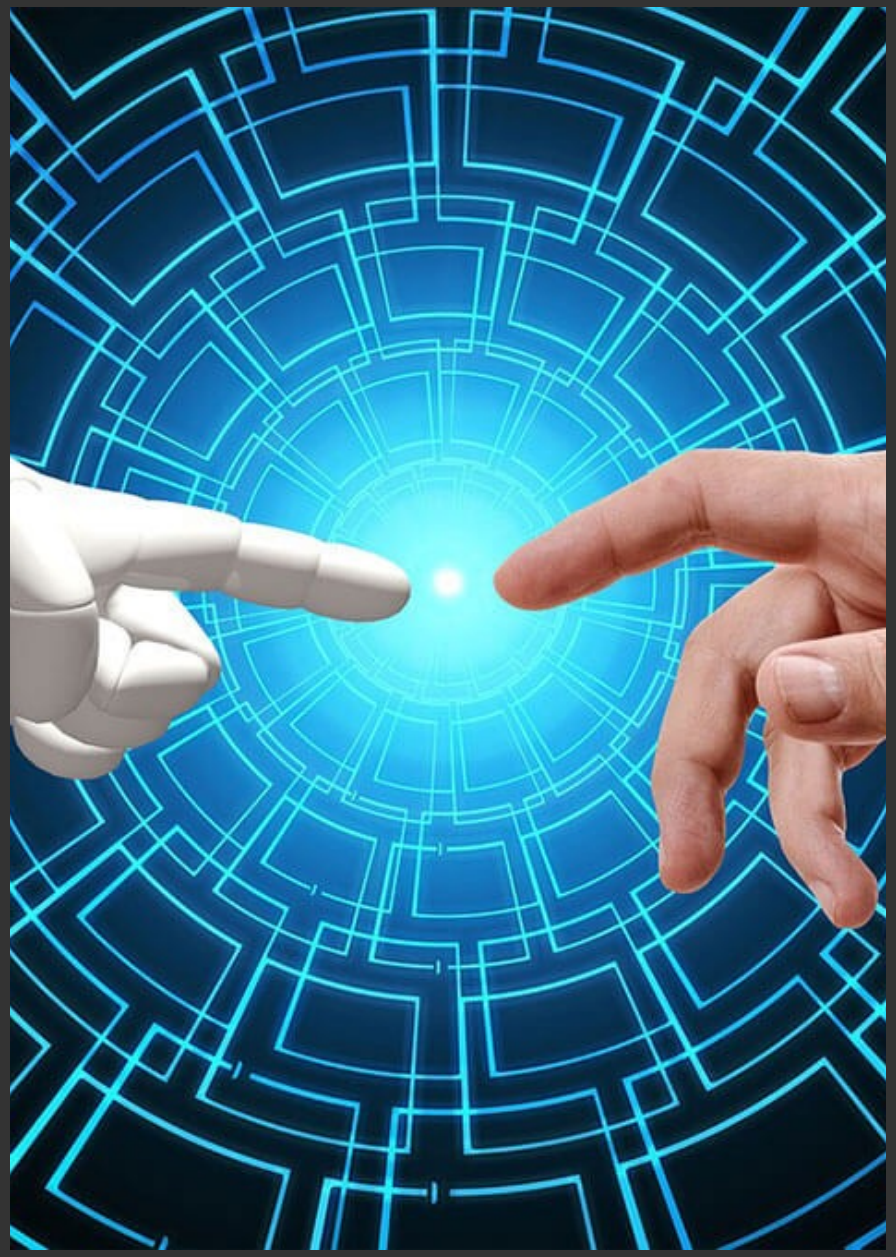
Matthew E. Taylor

PhD in AI, 2008

University of Alberta
Alberta Machine Intelligence Institute
AI Redefined

Francois Chabot

Co-Founder, Technology @ AIR



ChatGPT: Generative AI

ChatGPT is a revolutionary AI technology that is making headlines around the world. It uses RLHF (Reinforcement Learning from Human Feedback) to lean. But its real power is from human-AI collaboration.

Human-In-the-Loop Learning: What and Why?

1. Data during development

Humans and their data can be used to provide feedback and insights before deployment.

2. Offline updates

Data collected from humans can be used to occasionally update models and algorithms offline, allowing for more accurate predictions.

3. Online updates

Data can be used to update models and algorithms online, allowing for more accurate predictions in real-time.

4. Human-Agent teaming

Human-agent teaming allows for a variety of different inputs, rather than unidirectional data flows, allowing for better predictions and true teaming.

Gaps in Common HILL Approaches

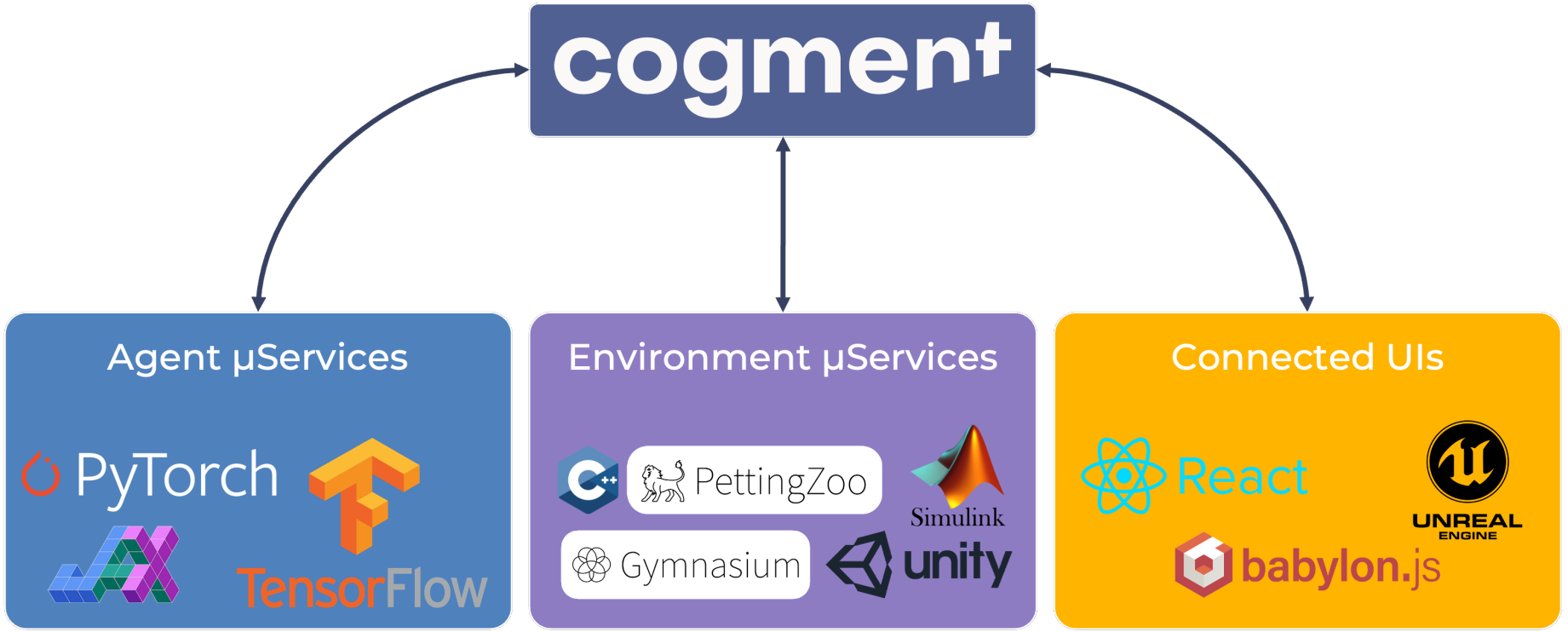
Adaptive

Trustworthy & Explainable

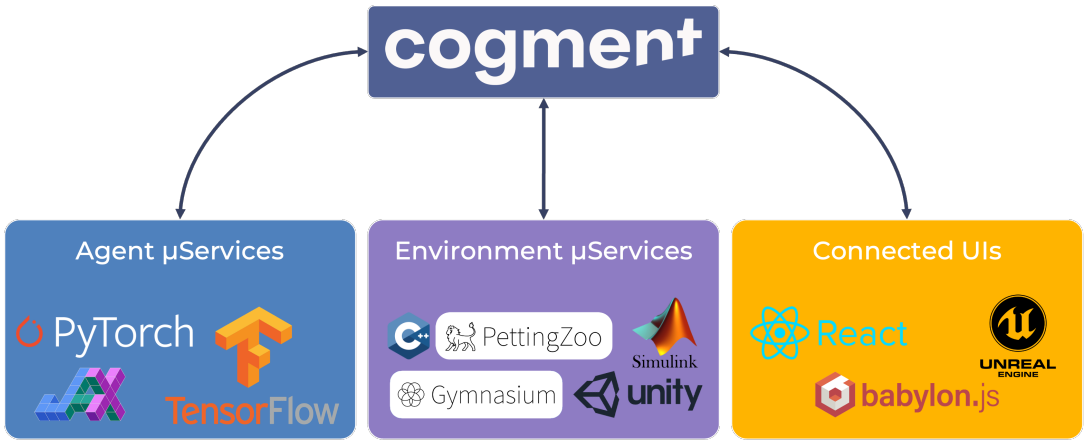
Human-centric, bidirectional teaming

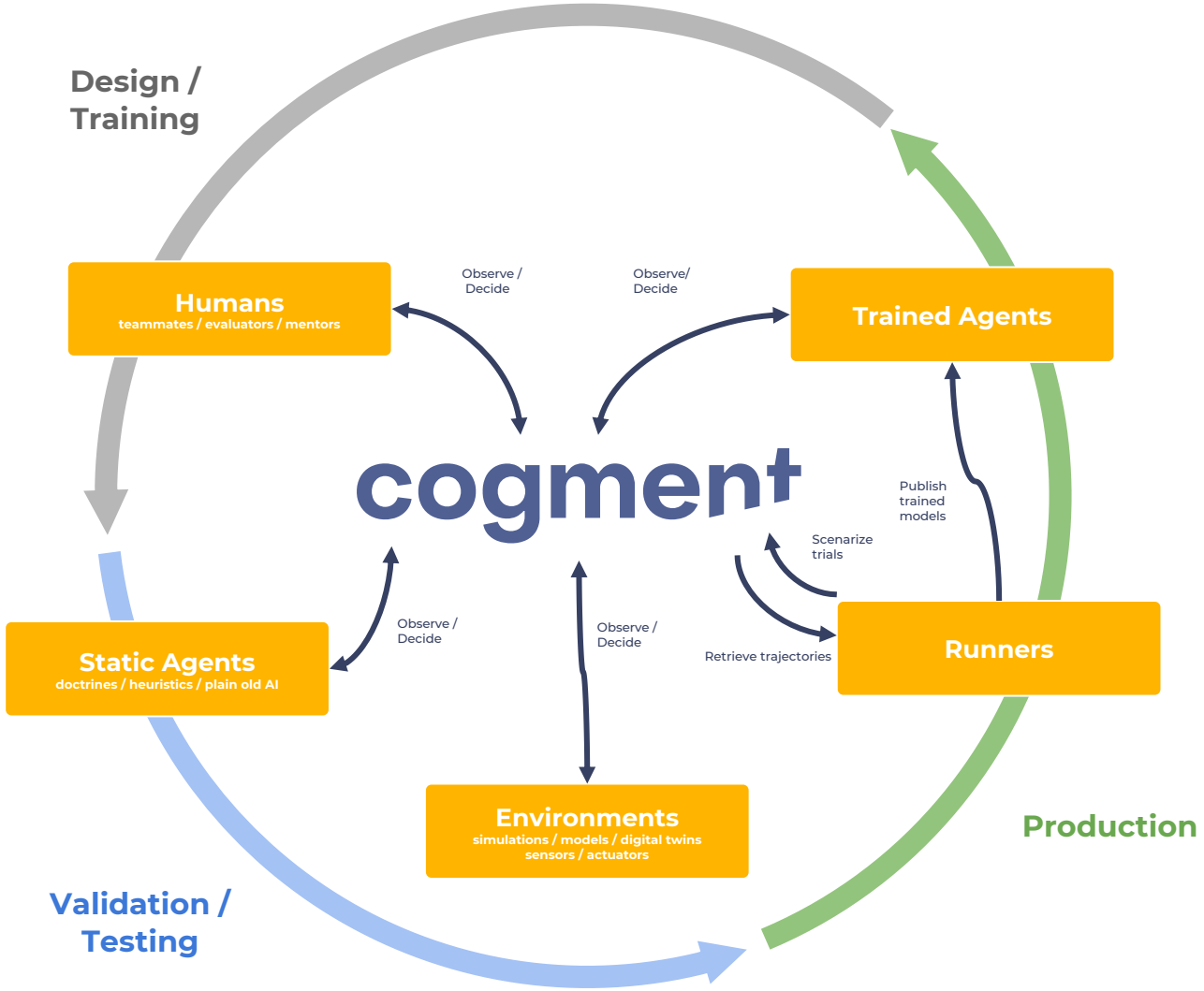
How can we best identify and address gaps in HILL approaches to ensure successful implementation and long-term sustainability?

Cogment: Interoperable Micro-Service Architecture



Cogment: Interoperable Micro-Service Architecture





Airfield Protection

Learning from:

- Environment
- Humans
- Each other

air
AI Redefined

JACOBB

THALES

UNIVERSITY OF ALBERTA

Airfield Protection

Learning from:

- Environment
- Humans
- Each other

air
AI Redefined

JACOBB

THALES

UNIVERSITY OF ALBERTA

Airfield Protection

Learning from:

- Environment
- Humans
- Each other

The Potential Risks of HILL

Data Scarcity

Anticipating & Understanding Humans

Novel Interaction Paradigms

Quick and Dirty is Difficult

The Potential Risks of HILL

- Data Scarcity
- Anticipating & Understanding Humans
- Novel Interaction Paradigms
- Quick and Dirty is Difficult

Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter

Attempt to engage millennials with artificial intelligence backfires hours after launch, with TayTweets account citing Hitler and supporting Donald Trump

https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter

The Potential ~~Risks~~ Benefits of HILL

Data Scarcity	Simulation & Pre-training
Anticipating & Understanding Humans	UX design, guardrails
Novel Interaction Paradigms	Human-centered frameworks
Quick and Dirty is Difficult	Engineering-focused tools

air
AI Redefined

The Intelligent Robot Learning Laboratory

The Incredible Benefits of HILL

Novel paradigms

HILL provides a new way of thinking about problem solving, allowing for more efficient and effective solutions.

Outperforming humans or AIs alone

HILL's combination of human and AI capabilities allows it to outperform either one alone.

Framework for teamwork

HILL provides a framework for collaboration between humans and AI, allowing for better problem solving.

Stop by Booth A60!