Human-Centered Artificial Intelligence

Fei-Fei Li

Sequoia Professor of Computer Science, Stanford University
Co-Director, Stanford University Institute for Human-Centered AI (HAI)

AI has risen with dizzying speed in recent years.

1987

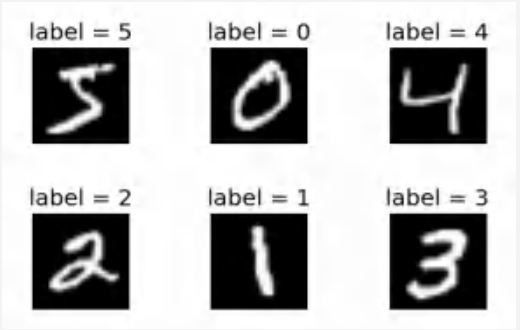
D. G. Lowe, *Artificial Intelligence,* 1987

1987

D. G. Lowe, *Artificial Intelligence,* 1987

1999

D. G. Lowe, *ICCV,* 1999



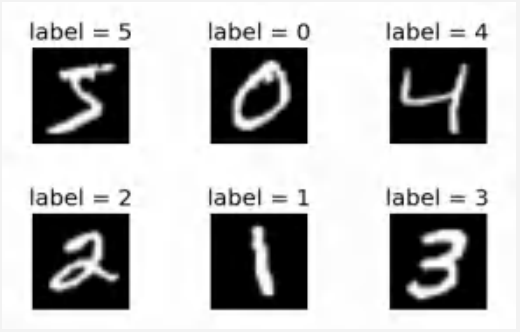
Y. LeCun, *Proc. IEEE,* 1998

1987

D. G. Lowe, *Artificial Intelligence,* 1987

1999

D. G. Lowe, *ICCV,* 1999



Y. LeCun, *Proc. IEEE,* 1998

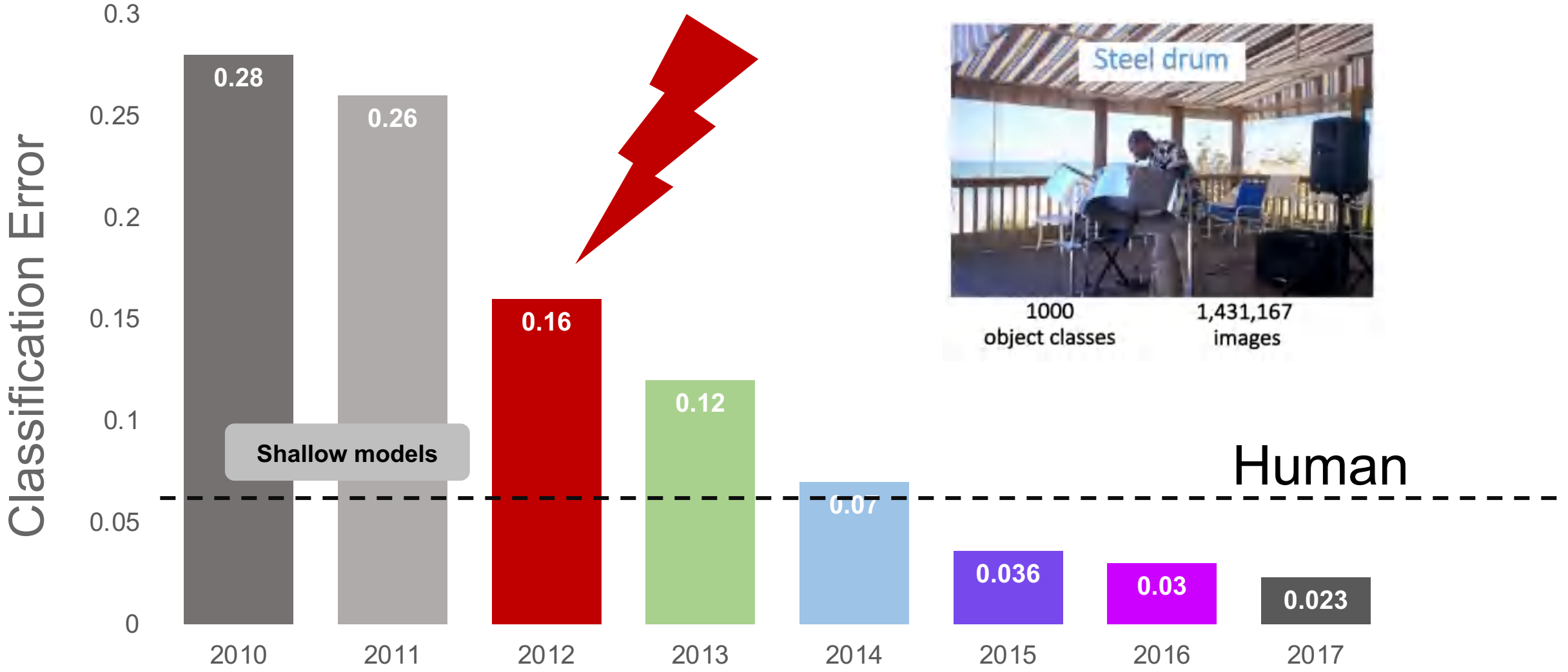
2012

"Leopard"

"Container Ship"

A. Krizhevsky, I. Sutskever & G. E. Hinton, *NIPS*, 2012
J. Deng, L. Fei-Fei et al. ImageNet, 2009

IMAGENET Challenge: Classification of 1000 Objects



Deng, J. et al. Fei-Fei, L. *CVPR*, 2009; Russakovsky et al. Fei-Fei, J. *IJCV*, 2012;

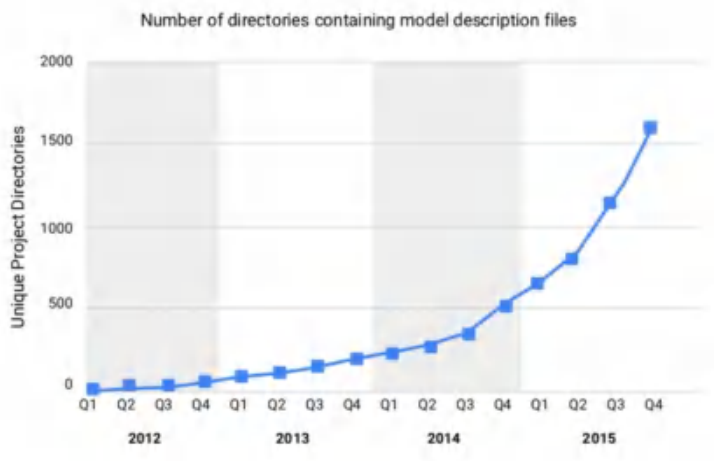
The Deep Learning Revolution

Computation

Algorithms

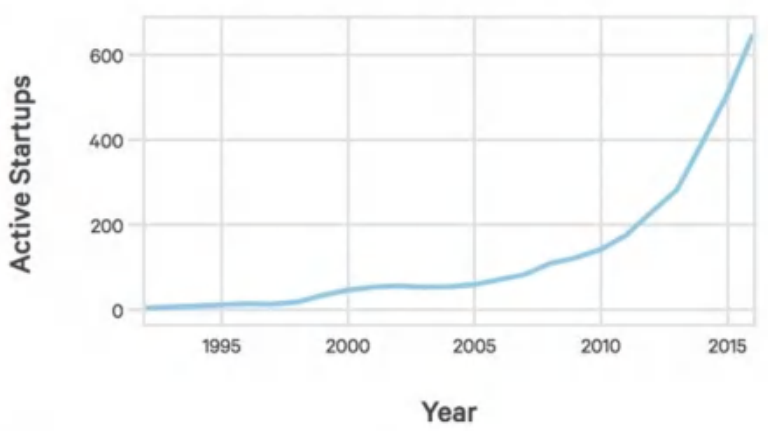
Big Data

AI’s Explosive Impact on Industry



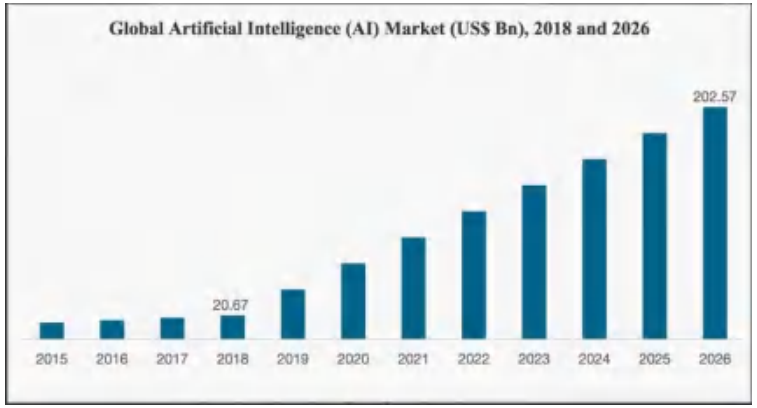
Growing Use of Deep Learning at Google

Source: Google



Startups Developing AI Systems

Source: Crunchbase, VentureSource, Sand Hill Econometrics



Global AI Market 2016-2026

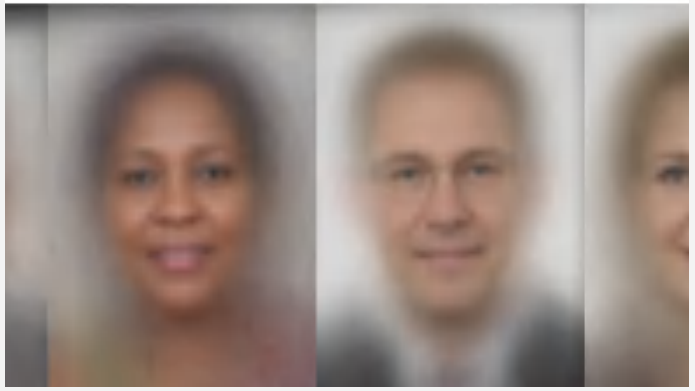
Source: fortunebusinessinsights.com


Google
self-driving car


Thank you, Google!
You don’t have to say that, dad.

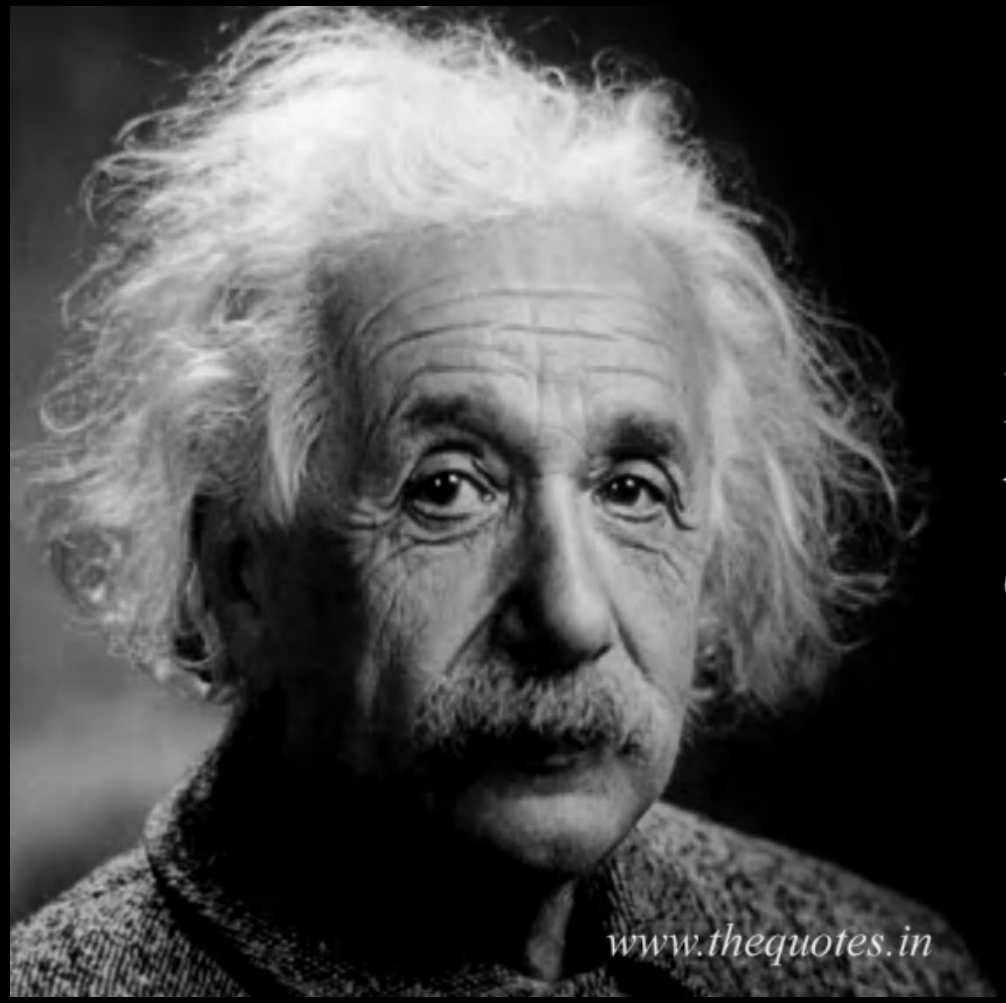
Job Displacement

Darker Male	Darker Female	Lighter Male	Lighter Female
94.0%	79.2%	100%	98.3%
99.3%	65.5%	99.2%	94.0%
88.0%	65.3%	99.7%	92.9%



Bias

Privacy



It has become appallingly obvious that our technology has exceeded our humanity.

Albert Einstein

A new approach to AI:

“Human-Centered AI”

Fei-Fei Li, How to make AI that’s good for people (op-ed), *New York Time,* 2018.

Human-Centered AI

H

The development of AI must be guided by a concern for its **human impact**.

A

AI should strive to **augment** and enhance us, not replace us.

I

AI must be more inspired by human **intelligence**.

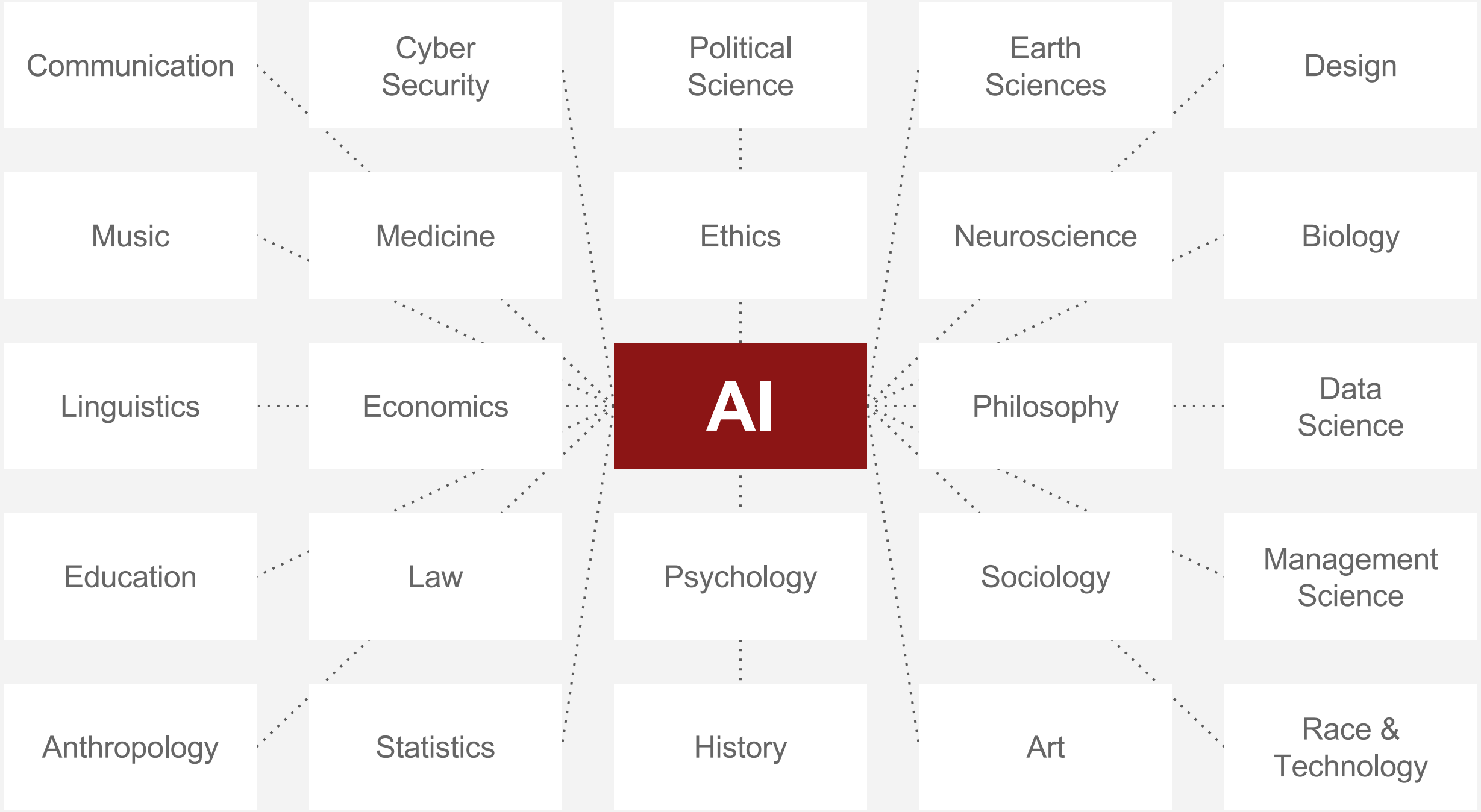
MARCH 18, 2019

Stanford University launches the Institute for Human-Centered Artificial Intelligence

The new institute will focus on guiding artificial intelligence to benefit humanity.

Human-Centered AI

The development of AI must be guided by a concern for its human **impact.**

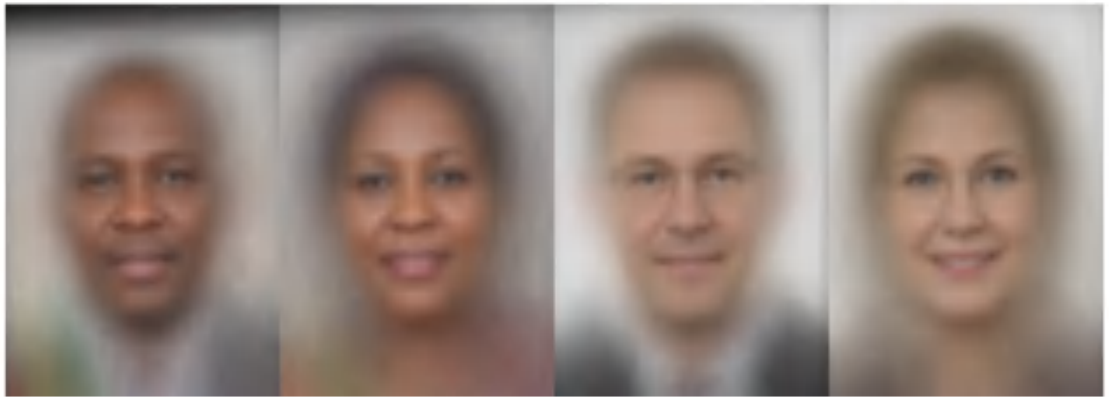

Communication
Cyber Security
Political Science
Earth Sciences
Design
Music
Medicine
Ethics
Neuroscience
Biology
Linguistics
Economics
AI
Philosophy
Data Science
Education
Law
Psychology
Sociology
Management Science
Anthropology
Statistics
History
Art
Race & Technology

Stanford HAI

Case Study

From ML Bias To ML Fairness

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%



J. Buolamwini & T. Gebru, *C. FAT*, PMLR*, 2018

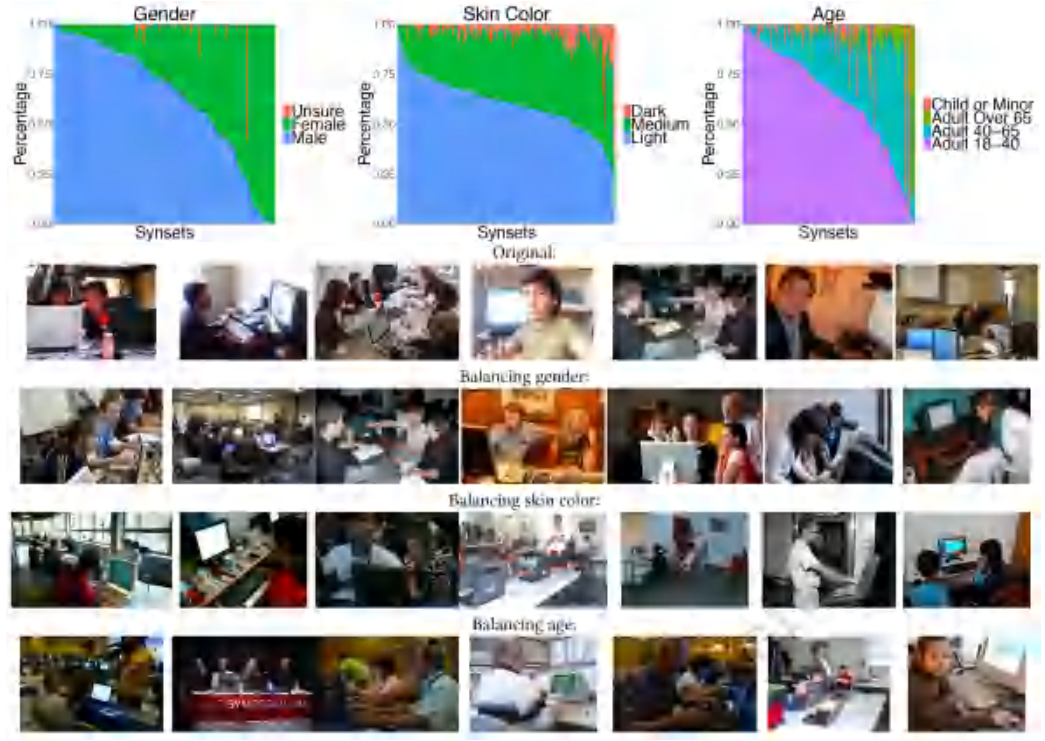
ML Fairness



Prof. Jia Deng &
Prof. Olga Russakovsky
(CS, Princeton)
Prof. Li Fei-Fei
(CS, Stanford)

- Dataset fairness
  (diversity and representation)

ImageNet Re-balancing (example: “programmer”)



K. Yang, et al. *C. FAT**, submitted

ML Fairness

Prof. James Zou (EE)
Prof. Londa Schiebinger (History)

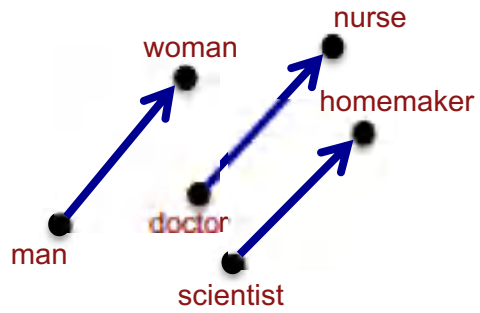
- Dataset fairness (diversity and representation)
- Algorithmic fairness (de-biasing and bias mitigation)

Intelligent Machines

How to Fix Silicon Valley's Sexist Algorithms

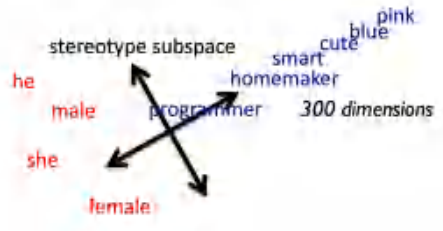
He's Brilliant, She's Lovely: Teaching Computers To Be Less Sexist

August 12, 2016 · 8:01 AM ET



Gender bias in word embedding

our solution



Algorithmic debiasing via geometry

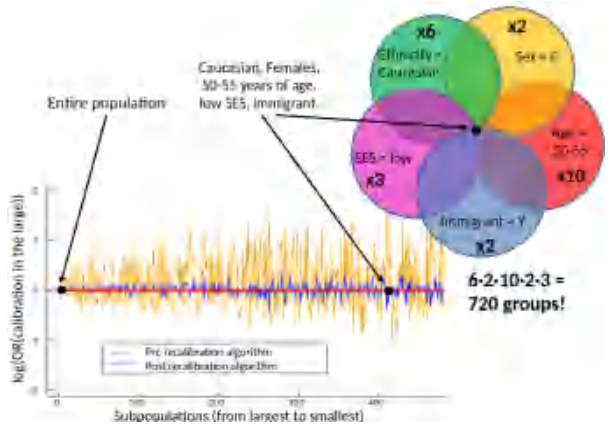
Ghorbani & Zou, ICML 2019; Tannenbaum, et al. Nature 2019
Zou & Schiebinger. Nature 2018; Garg et al. PNAS, 2018.

ML Fairness

Prof. Omer Reingold (CS)

- Dataset fairness (diversity and representation)
- Algorithmic fairness (de-biasing and bias mitigation)
- Computing fairness (theoretical guarantees)

- **New definition of fairness** -> Aim to reduce discrimination for every large subgroup that can be identified computationally
- An iterative **"fairness" algorithm** to constantly "nudge" subpopulation averages to their correct values



Cardiology Application:
N. Barda, N. Dagan, 2019

Calibration in the large:
1.36 → 1.02

Decreased variance between sub-groups:
96.2%

O. Reingold, FOCS, 2019; S. Garg, et al. EC, 2019; U. Hebert-Johnson et al., ICML, 2018; M. Kim, CoRR, 2018; C. Dwork et al. ITCS, 2012

Stanford HAI

ML Fairness

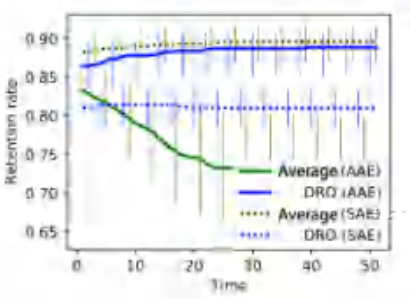


Prof. Tatsu Hashimoto
& Prof. Percy Liang (CS)
Prof. John Duchi
(Statistics & EE)

- Dataset fairness
  (diversity and representation)
- Algorithmic fairness
  (de-biasing and bias mitigation)
- **Computing fairness**
  (theoretical guarantees)

Preventing disparity amplification

Distributional robustness
Protects minority
performance over time



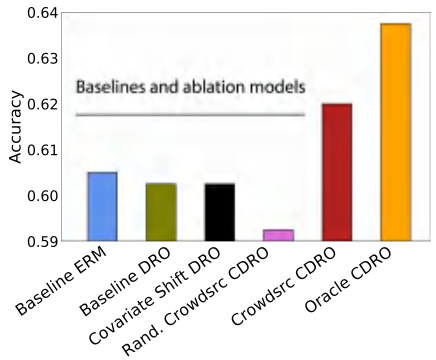
Hashimoto, Srivastava, Namkoong, Liang
ICML 2018 (Best paper runner up)

Human counterfactual robustness

Problem:
Spurious correlations

Race ⟷ Crime
Income

Solution:
Robustness w/ Human counterfactuals



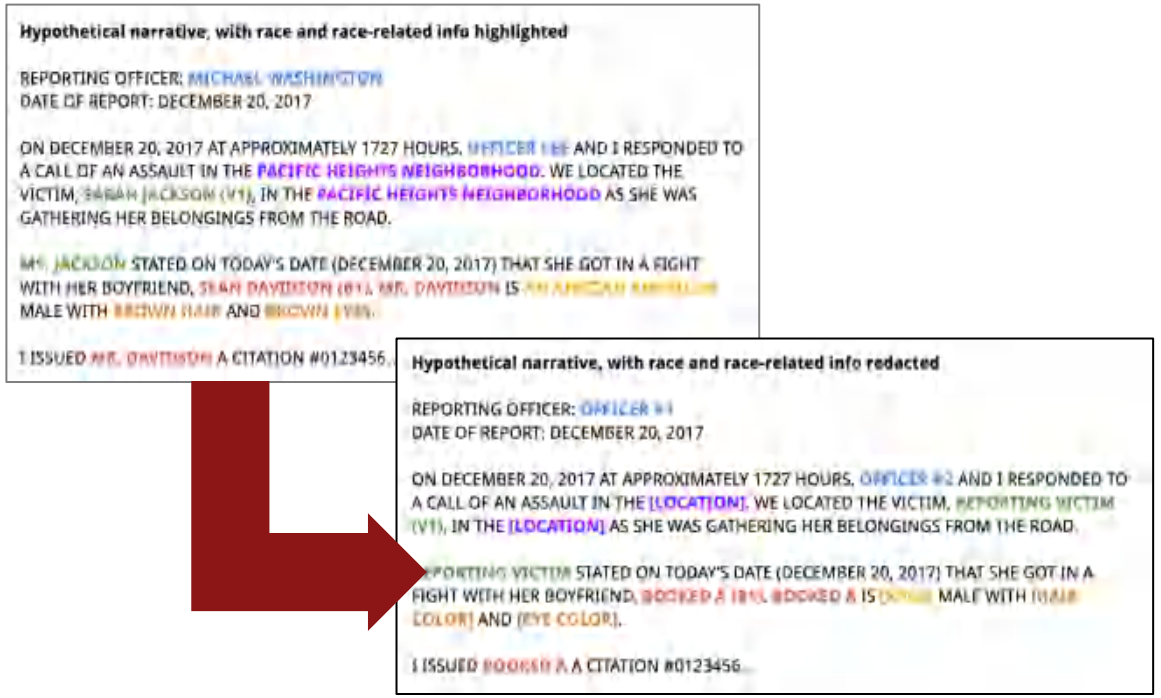
Srivastava, Hashimoto, Liang
(under review) AISTATS

ML Fairness

Prof. Sharad Goel
(Management Sci & Eng.)

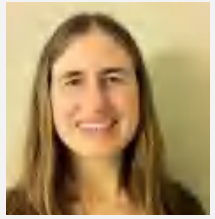
- Dataset fairness
  (diversity and representation)
- Algorithmic fairness
  (de-biasing and bias mitigation)
- Computing fairness
  (theoretical guarantees)
- **Decision-making fairness**
  (race-blind decisions)

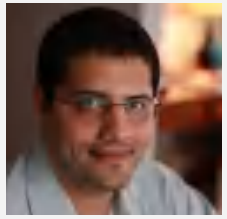
Enabling race-blind decisions
Stanford Computational Policy Lab



Corbett-Davies, S., & Goel, S. (2018). arXiv:1808.00023.
Lin, Z., Chohlas-Wood, A., & Goel, S. (2019). Conference on AI, Ethics, and Society ACM.

ML Fairness

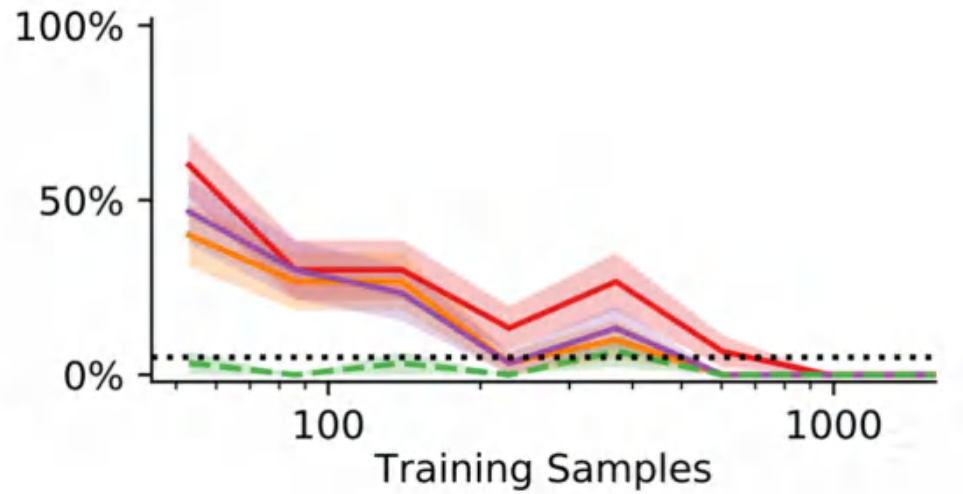




Prof. Emma Brunskill
(CS/Education, Stanford)
Prof. Phil Thomas
(CS, U. Mass. Amherst)

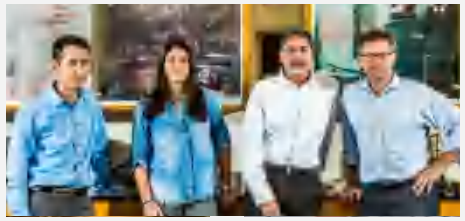
- Dataset fairness
  (diversity and representation)
- Algorithmic fairness
  (de-biasing and bias mitigation)
- Computing fairness
  (theoretical guarantees)
- **Decision-making fairness**
  (gender-blind decisions)

- Experiments where a simple tutoring system had multiple variants, some purposefully designed to be “unfair” showed other algorithms could be sexist
- High probability guarantee that resulting solution will satisfy desired fairness constraints: **Fair Batch Decision Making Under Uncertainty**



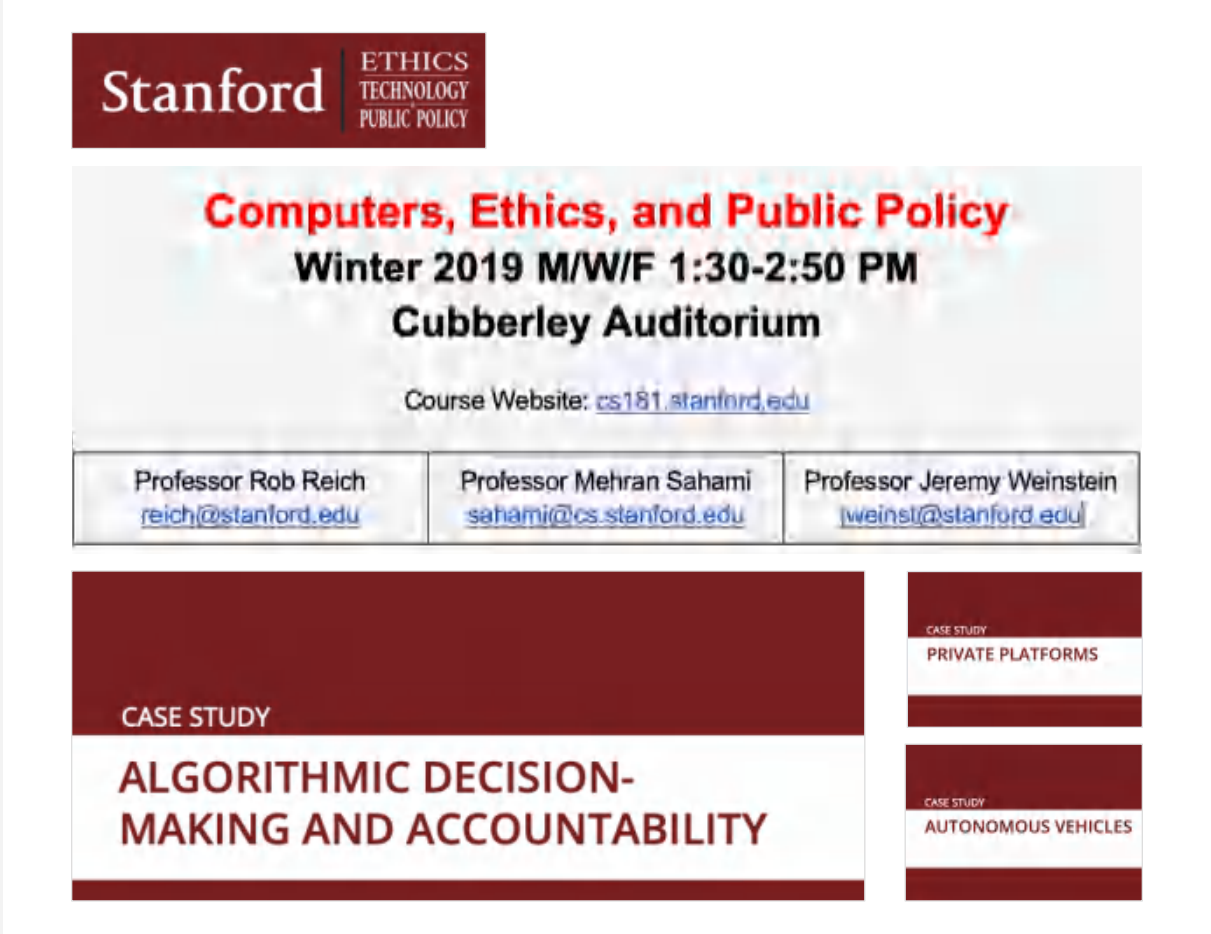
B. Metevier, et al. *NeurIPS 2019*

ML Fairness

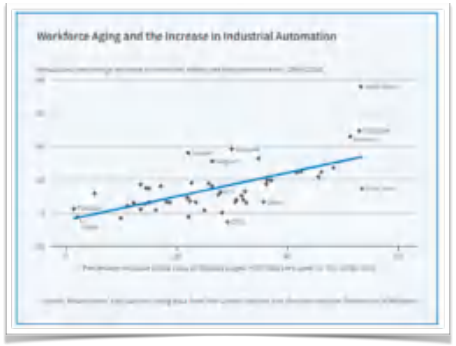


Prof. Rob Reich
(Political Science)

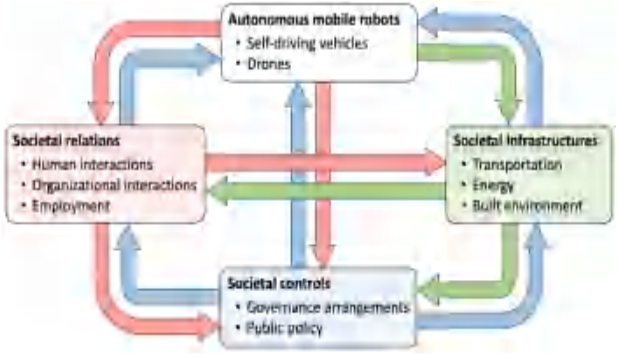
- Dataset fairness
  (diversity and representation)
- Algorithmic fairness
  (de-biasing and bias mitigation)
- Computing fairness
  (theoretical guarantees)
- Decision-making fairness
- **Ethics education**
  (multidisciplinary course)



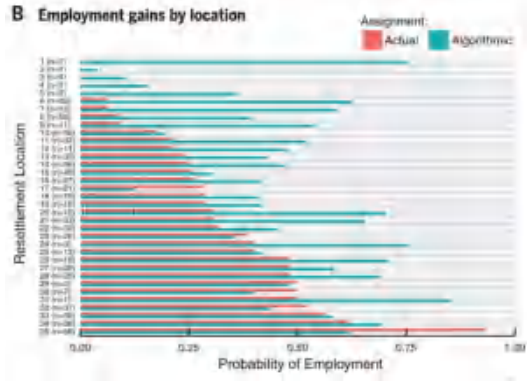
Stanford
HAI



Prof. Susan Athey (Economics)
AI + Future of Work



Profs. Marco Pavone (AA), David Grusky (Soc), Mark Duggan (Econ)
Societal Impact on Autonomous Robots



Prof. Jens Heinmueller (Pol. Sci.)
AI + Refugee Policy

Human-Centered AI

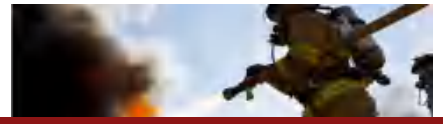
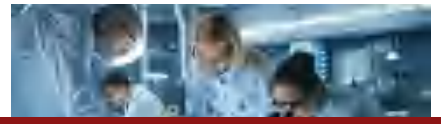
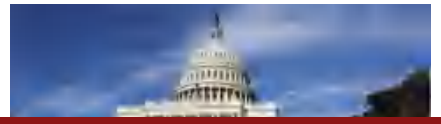
A

AI should strive to **augment** and enhance us, not replace us.





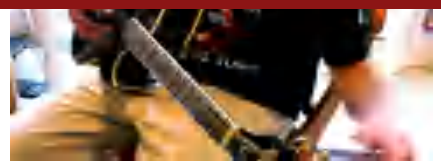


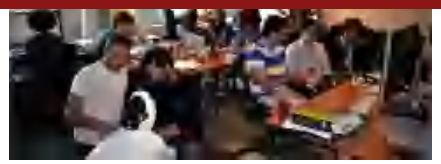

~50%

Of all current work activities can theoretically be automated now.

McKinsey, 2017

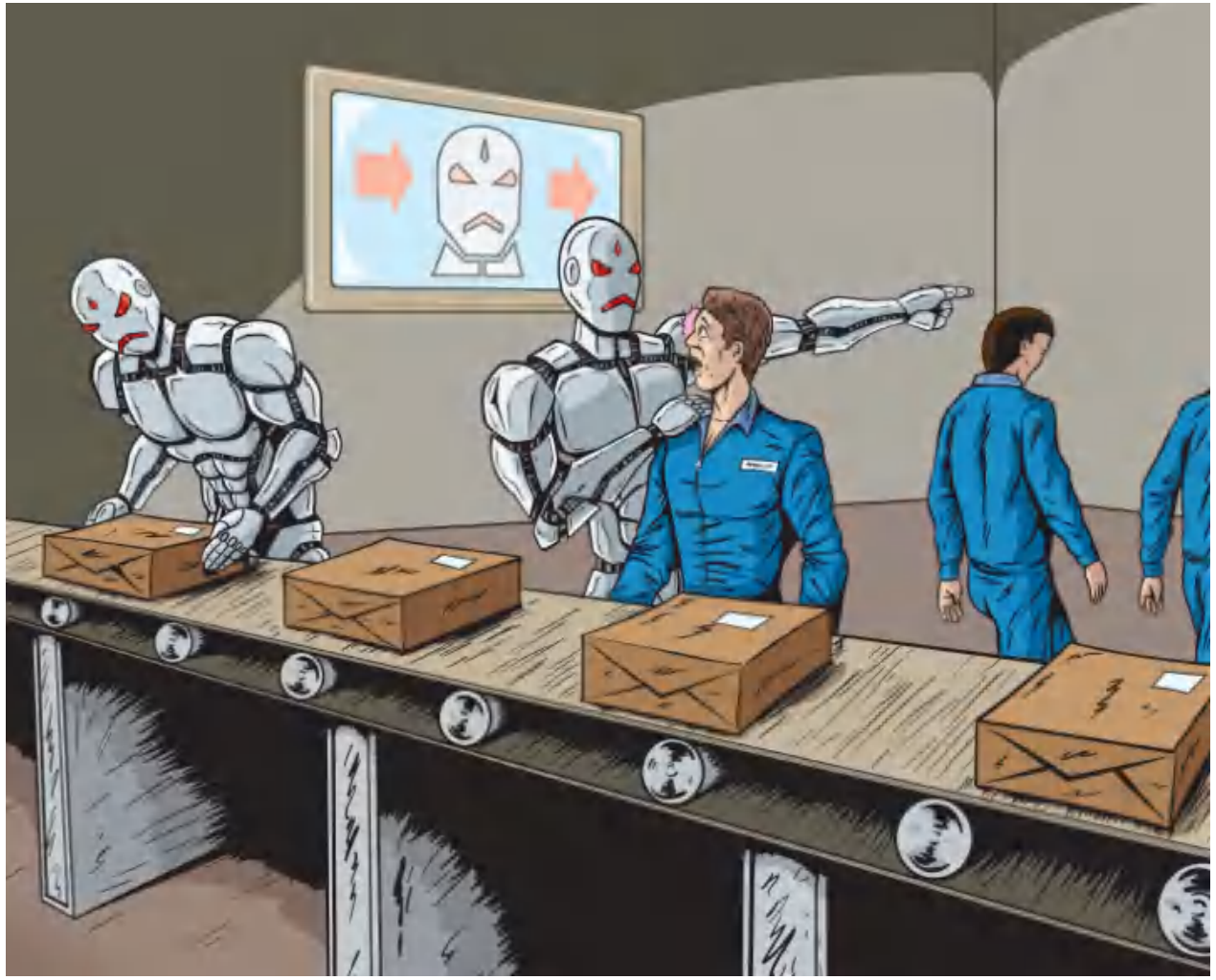









“Replace”



“Augment”

Case Study: AI-assisted Healthcare delivery

Collaborators

Computer Science

Fei-Fei Li

Alexandre Alahi

Gabriel Bianconi

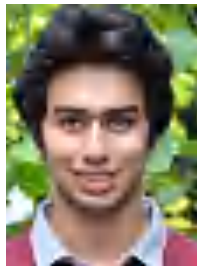
Michelle Guo

Albert Haque

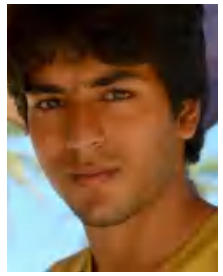
Tim Hsieh

Bingbin Liu

Zelun Luo



Rishab Mehra



Sanyam Mehra

Alisha Rege

Serena Yeung

Lucile Packard Children's Hospital Stanford

Stanford HEALTH CARE

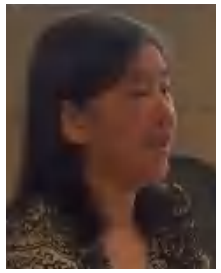
Intermountain Healthcare

ON LOK

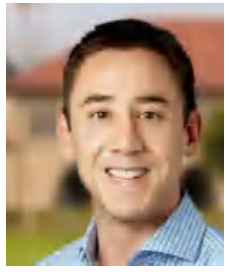
Medicine

Arnold Milstein

William Beninati



Wanda Chin

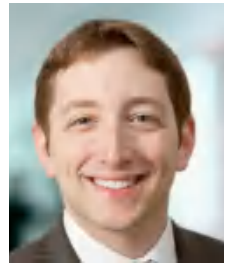


Lance Downing

Jeffrey Jopling

Grace Li

Jay Luxenberg



Terry Platchek

Amit Singh

Enhancing human care with intelligent systems

Yeung, Downing, Fei-Fei, Milstein. *New England Journal of Medicine (NEJM)*, 2018.

America's Medical Error

250K Deaths

Annually

Hospital-Acquired Infections

99K Deaths

Annually

Unmonitored Elderly Fall Injuries

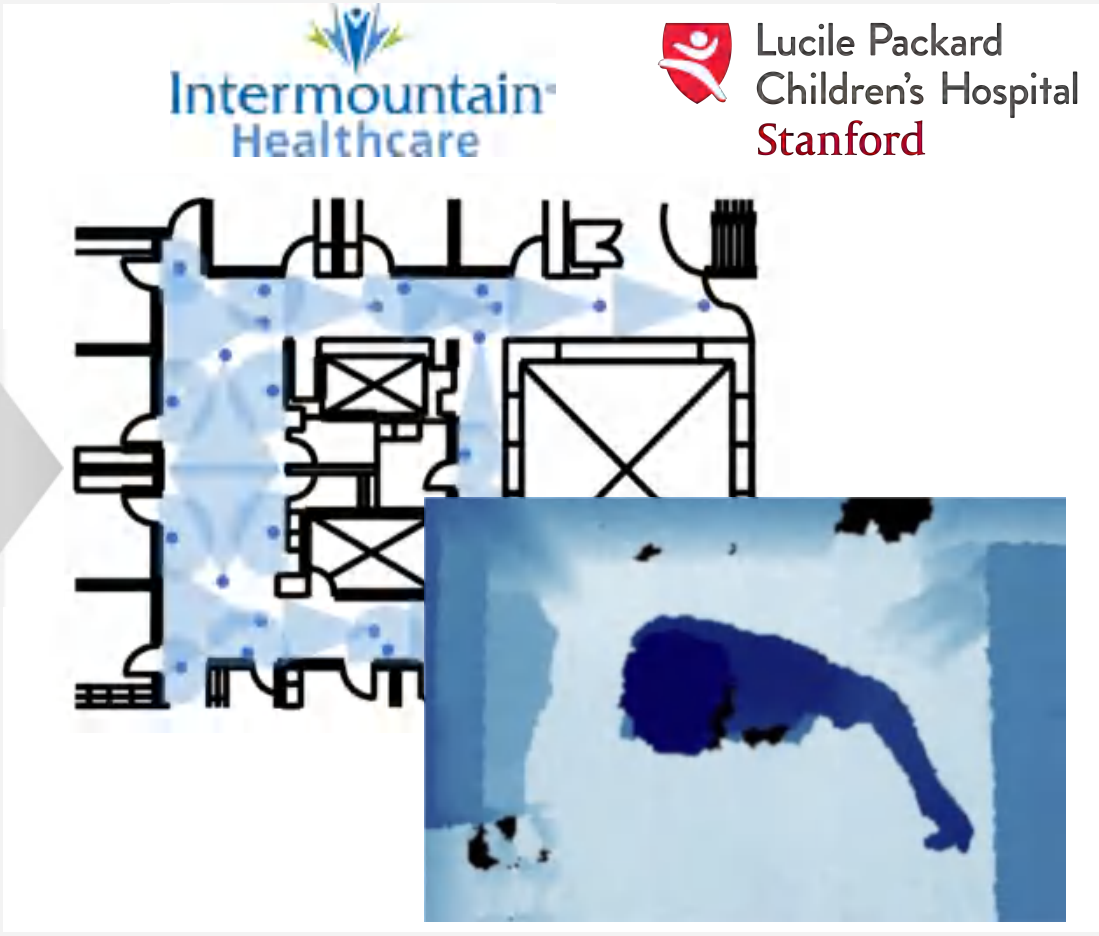
$36.4B

Annually

A. Houser, W. Fox-Grage & K. Ujvari, *AARP Public Policy Institute*, 2012)

Airtek Indoor Air Solutions. 2014. Calfee.

Annual Review of Medicine 2012

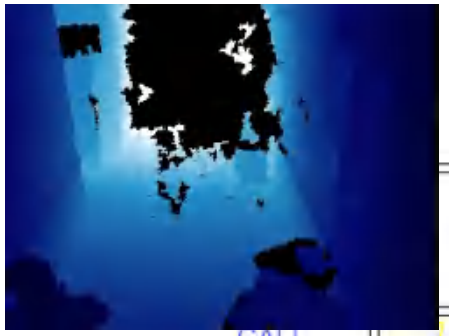


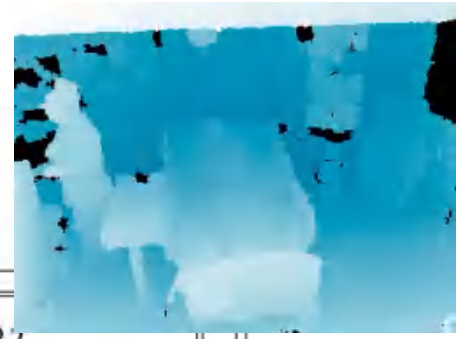
From: Inconsistent hand hygiene

To: Intelligent sensors placed throughout hospitals

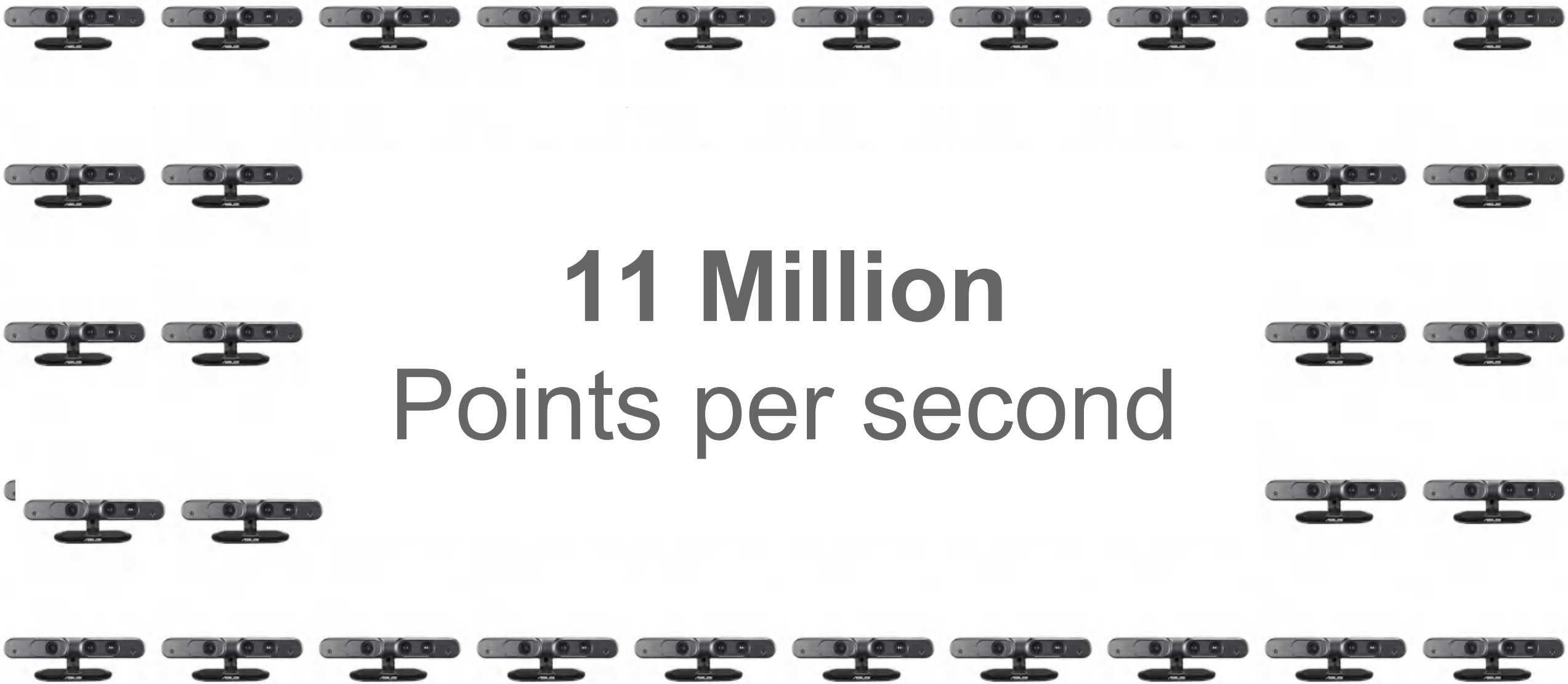
A. Haque, A. Singh, A. Alahi, S. Yeung, M. Guo, A. Luo, J. Jopling, L. Downing, W. Beninati, T, Platchek, A. Milstein & L. Fei-Fei, *Under review*
A. Haque, E. Peng, A. Luo, A. Alahi,S. Yeung & L. Fei-Fei, *ECCV*, 2016

Smart sensors throughout a hospital unit



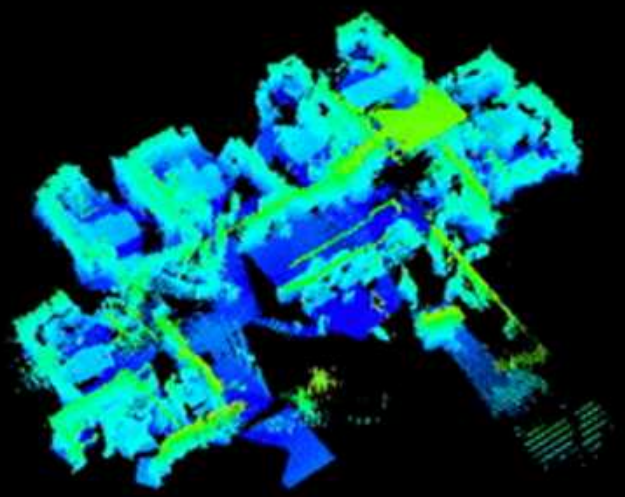


Guo, Haque, Deru, Campbell, Milstein, Beninati, Fei-Fei (2019). *In preparation*.

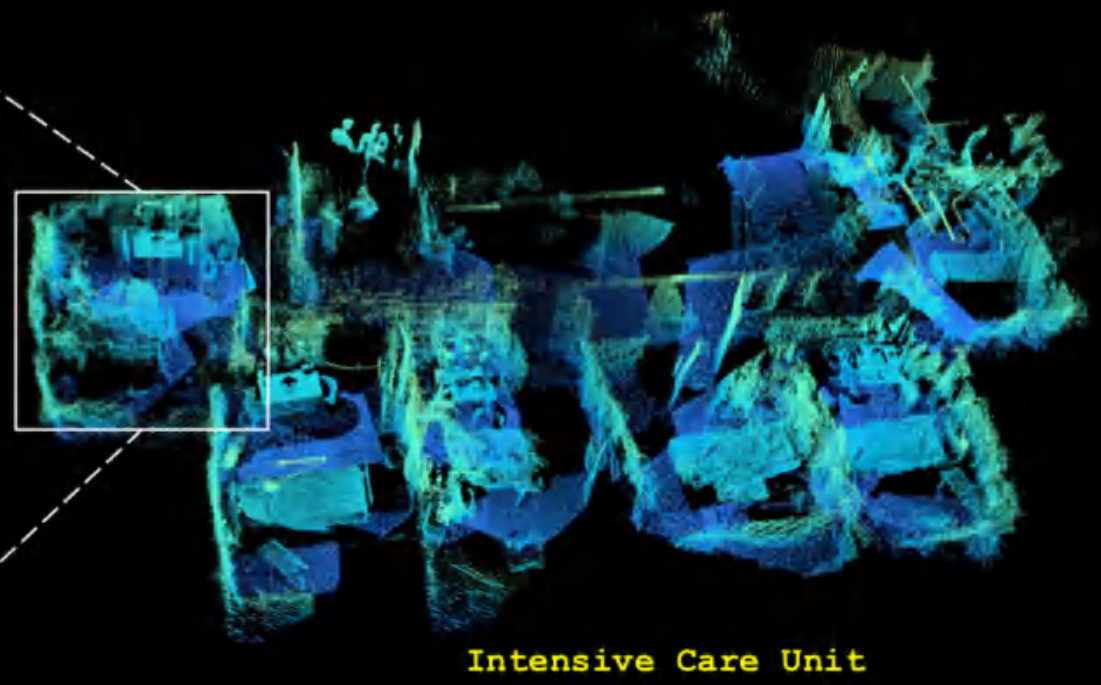


11 Million
Points per second

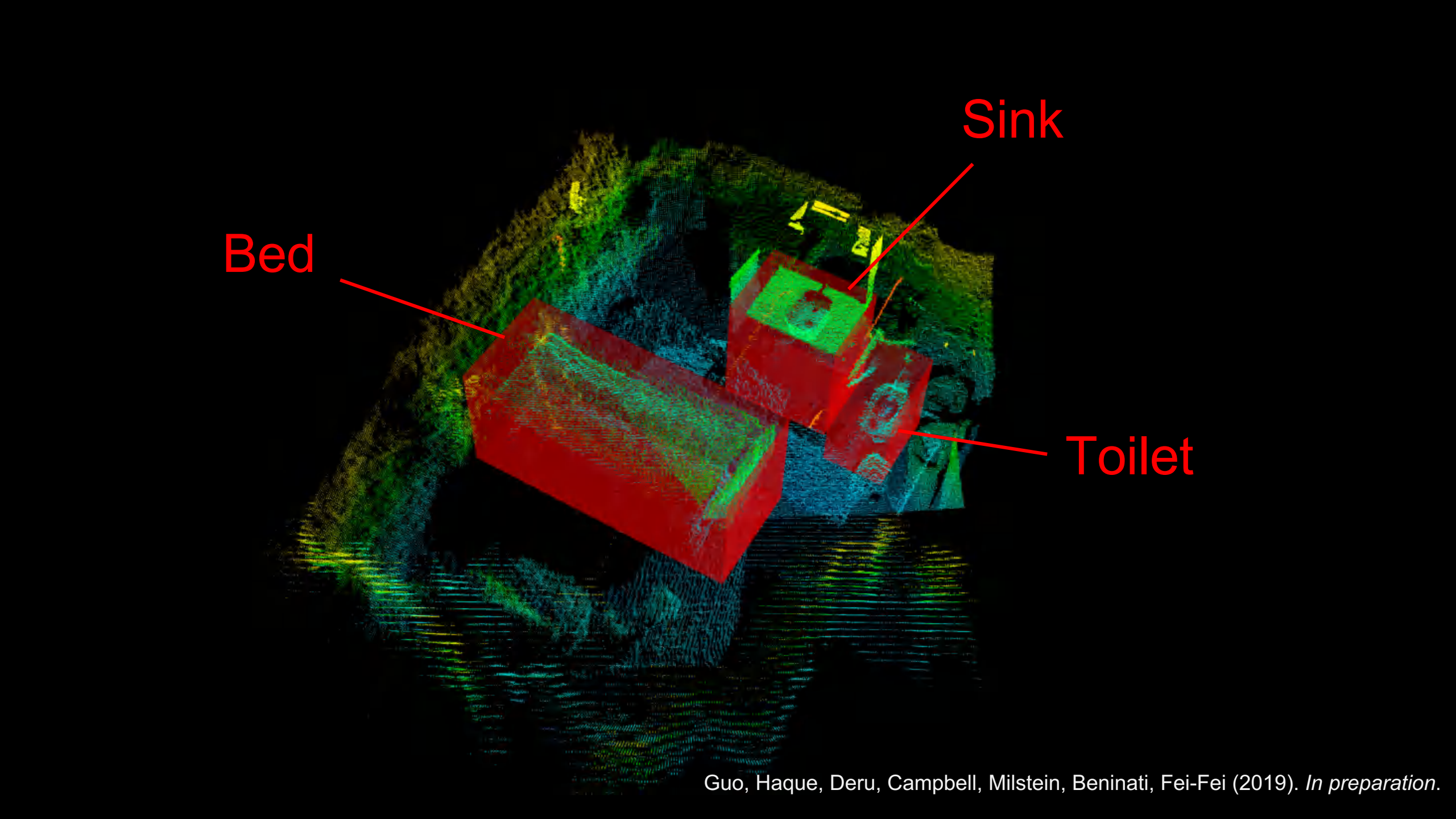
Guo, Haque, Deru, Campbell, Milstein, Beninati, Fei-Fei (2019). *In preparation*.



Guo, Haque, Deru, Campbell, Milstein, Beninati, Fei-Fei (2019). *In preparation*.



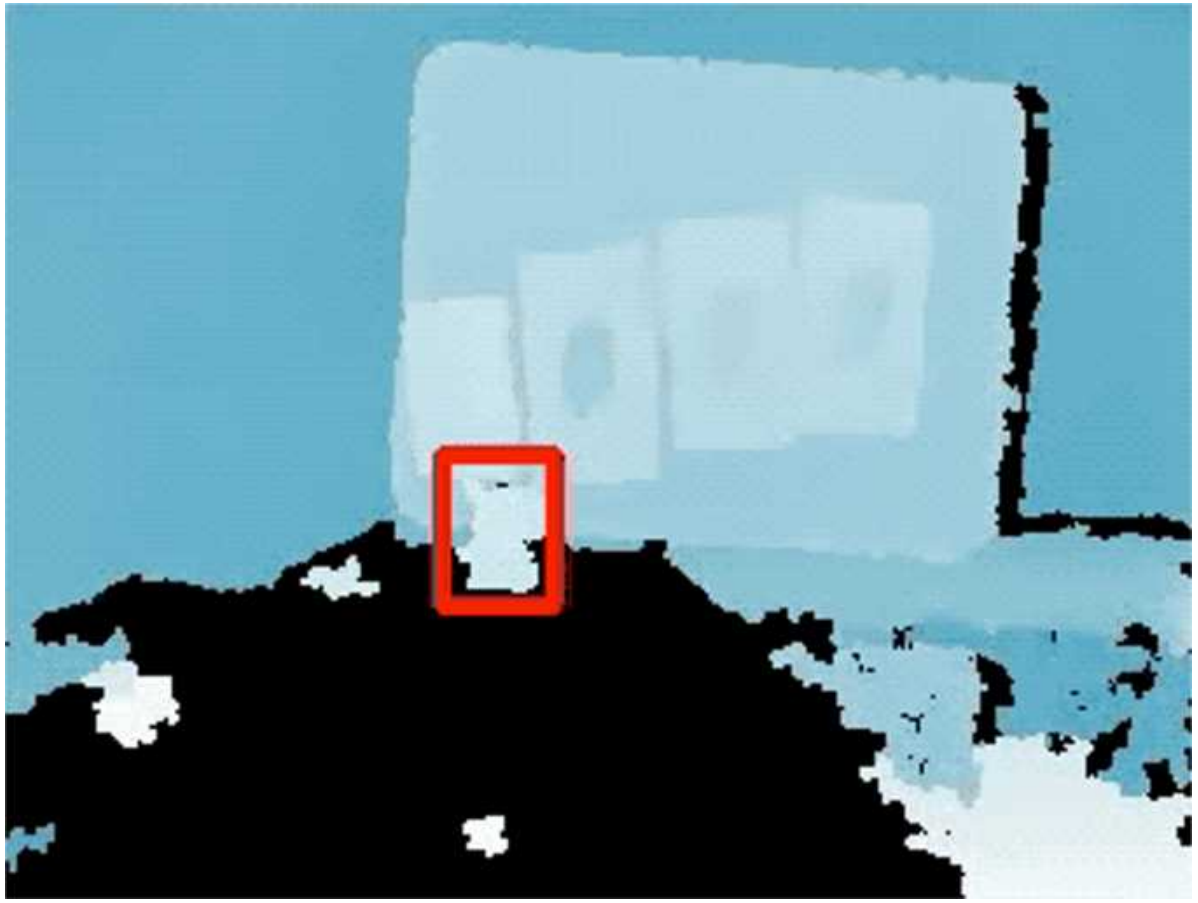
Guo, Haque, Deru, Campbell, Milstein, Beninati, Fei-Fei (2019). *In preparation*.



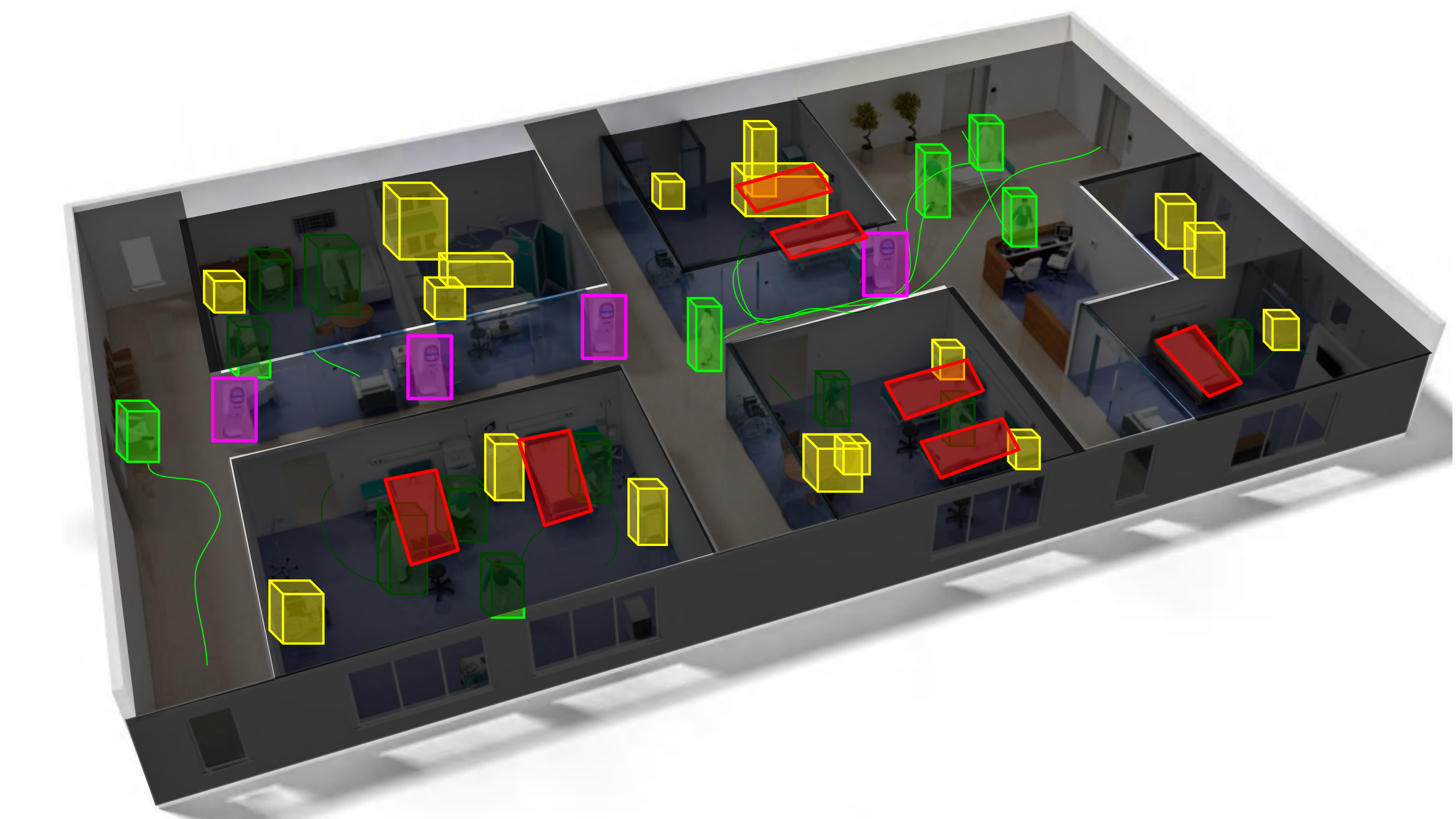
Guo, Haque, Deru, Campbell, Milstein, Beninati, Fei-Fei (2019). *In preparation*.

Clean Exit

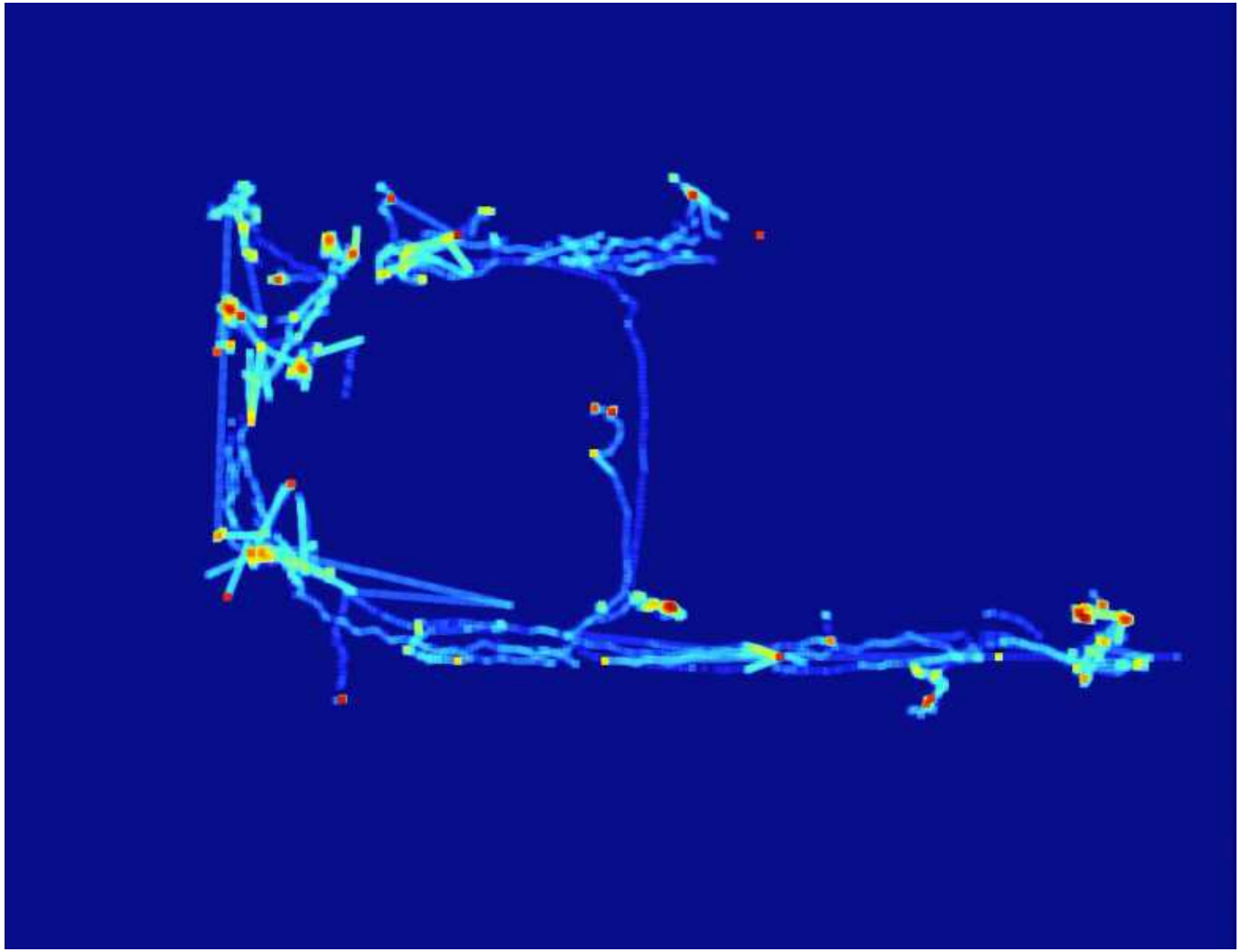
Dirty Exit



Guo, Haque, Deru, Campbell, Milstein, Beninati, Fei-Fei (2019). *In preparation*.

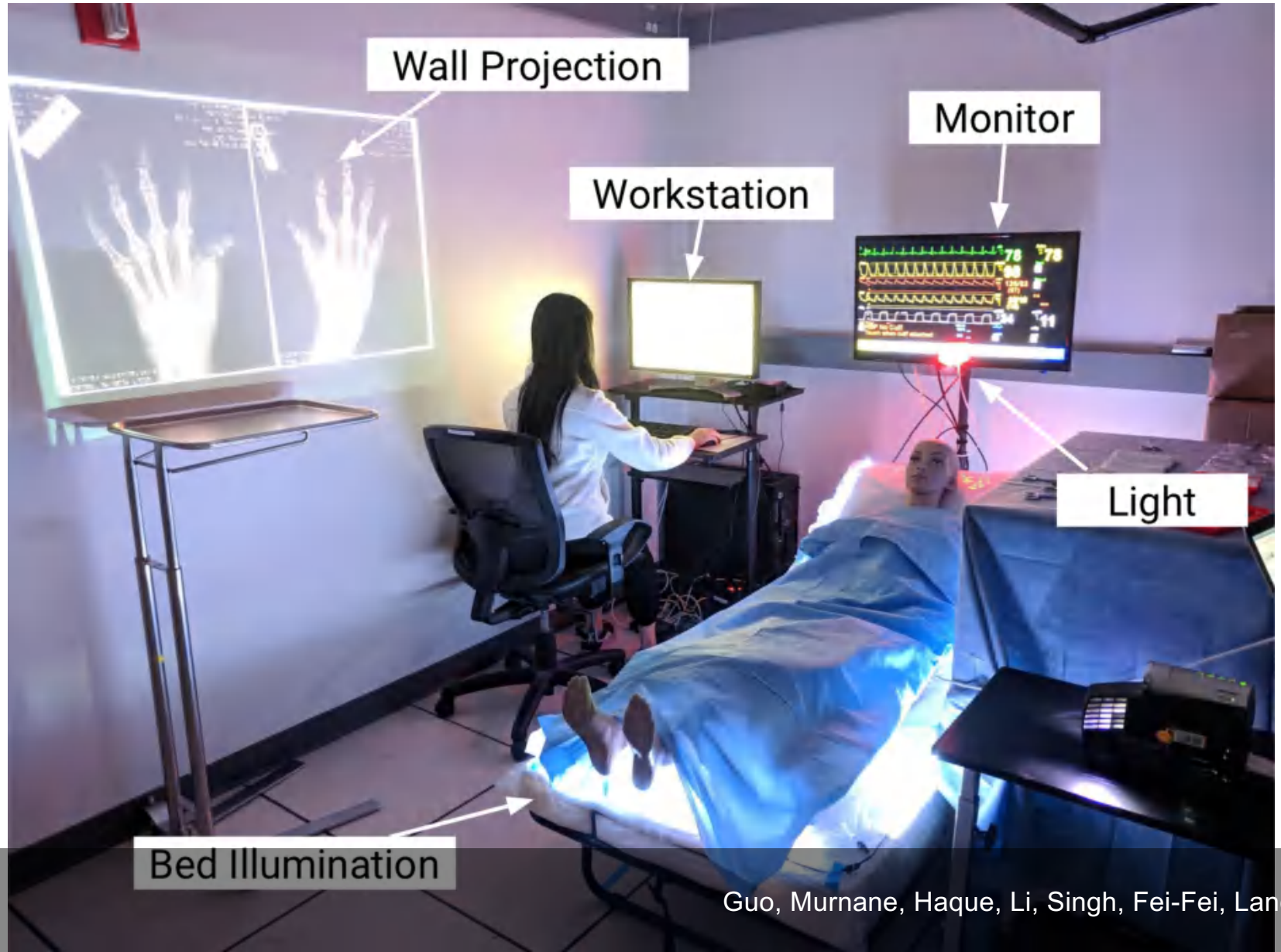


Aggregated tracks can be used for space analytics

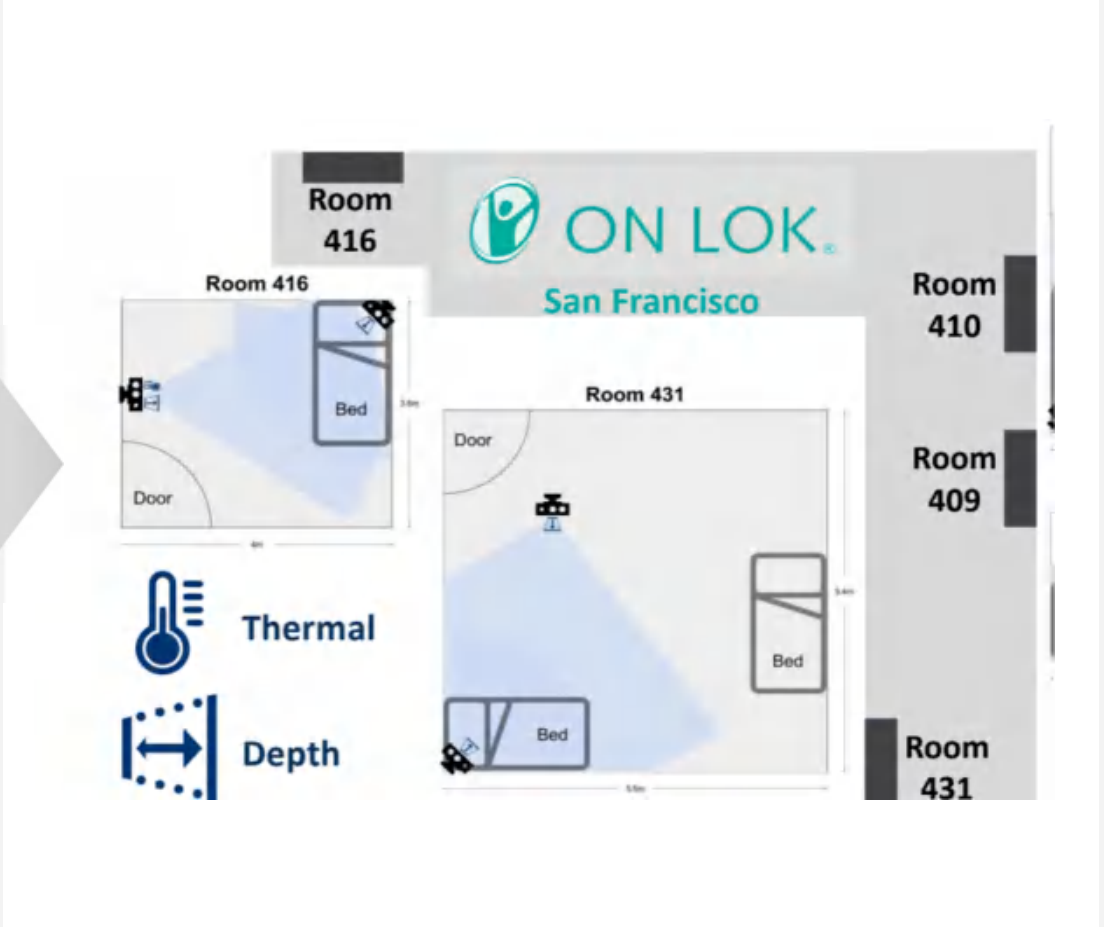


Haque, Guo, Alahi, Yeung, Luo, Rege, Singh, Jopling, Downing, Beninati, Platchek, Milstein & Fei-Fei. MLHC. 2017.

From observation to clinical behavioral changes



Guo, Murnane, Haque, Li, Singh, Fei-Fei, Landay (2019). *In preparation.*

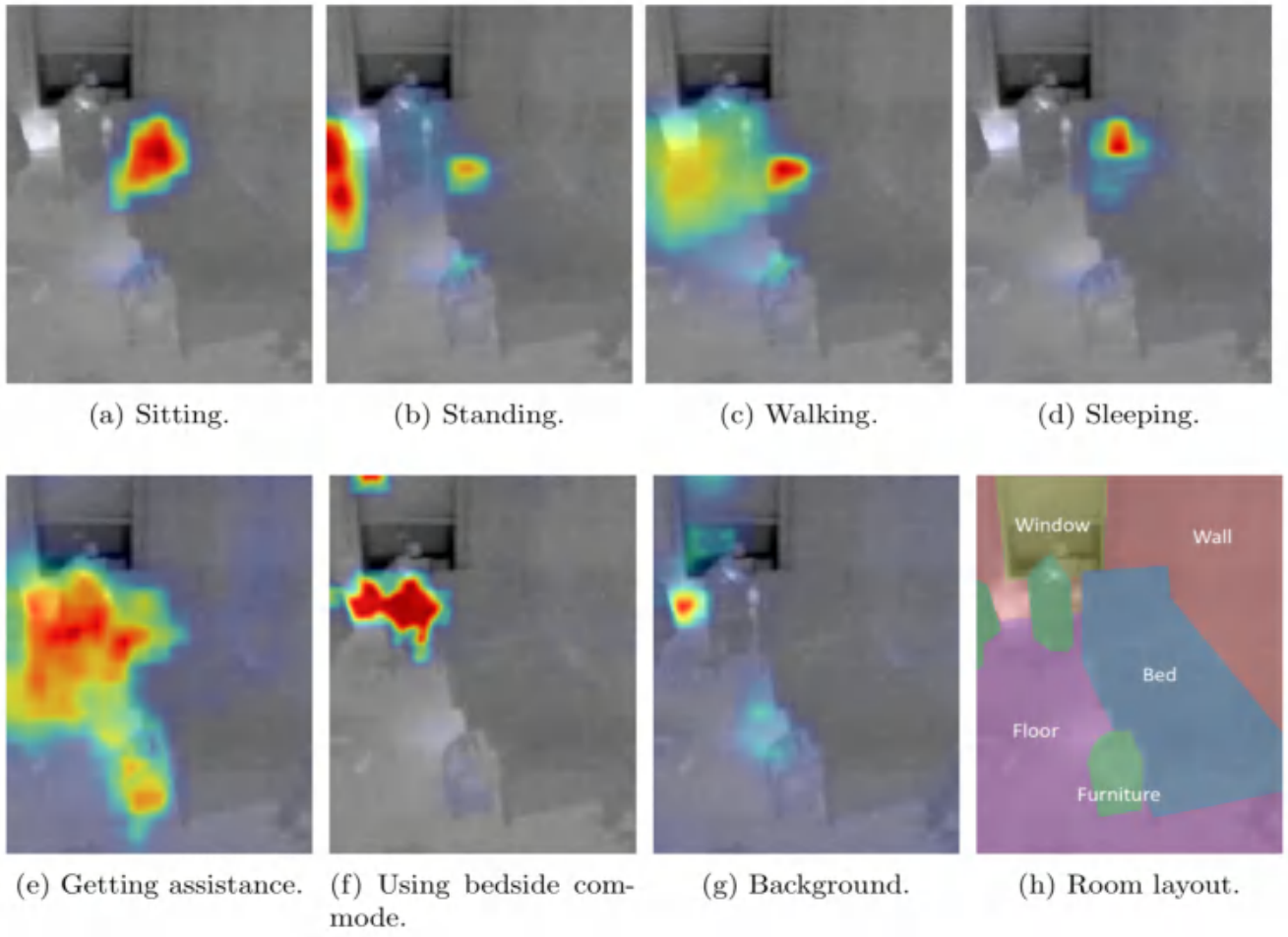


From: Ineffective wearables, lack of human caretakers

To: Intelligent sensors placed throughout senior living homes

A. Luo, T. Hsieh, R. Rege, A. Mehra,G. Pusiol, L. Downing, A. Milstein & L. Fei-Fei. *In preparation.*

Our system analyzes ***where*** each activity happens

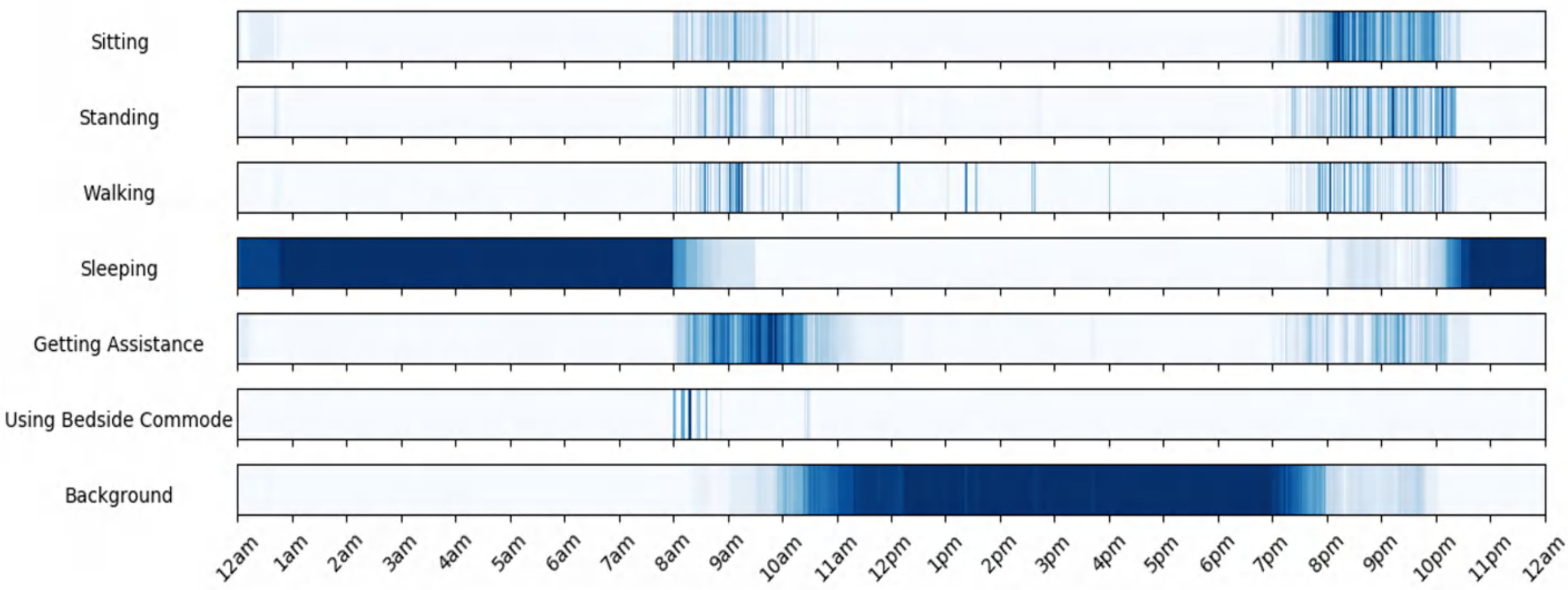


(a) Sitting. (b) Standing. (c) Walking. (d) Sleeping.

(e) Getting assistance. (f) Using bedside commode. (g) Background. (h) Room layout.

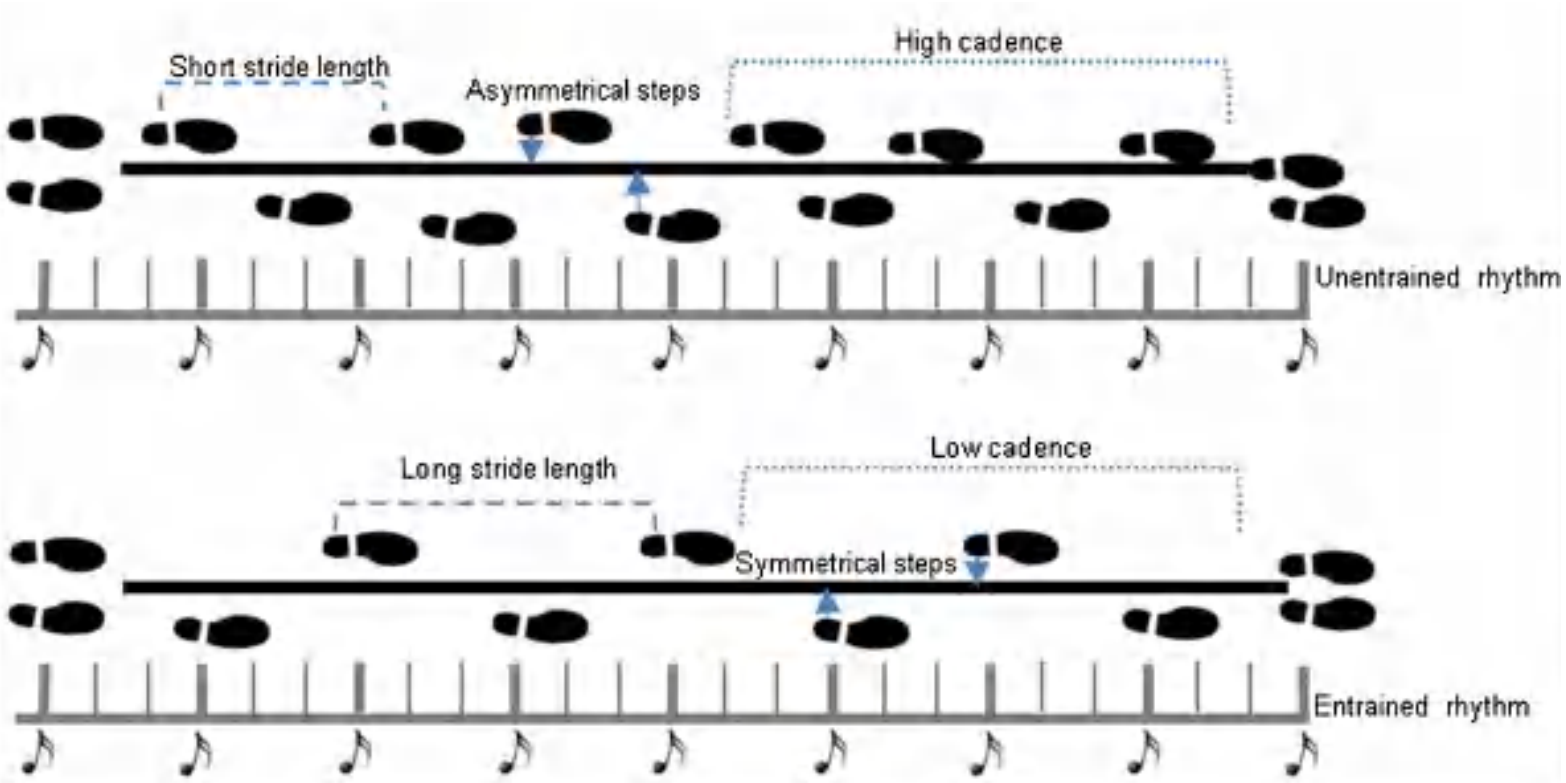
Luo*, Hsieh*, Balachandar, Yeung, Pusiol, Luxenberg, Li, Li-Jia, Downing, Milstein, Fei-Fei. MLHC 2018.

and ***when*** each activity happens



Luo*, Hsieh*, Balachandar, Yeung, Pusiol, Luxenberg, Li, Li-Jia, Downing, Milstein, Fei-Fei. MLHC 2018.

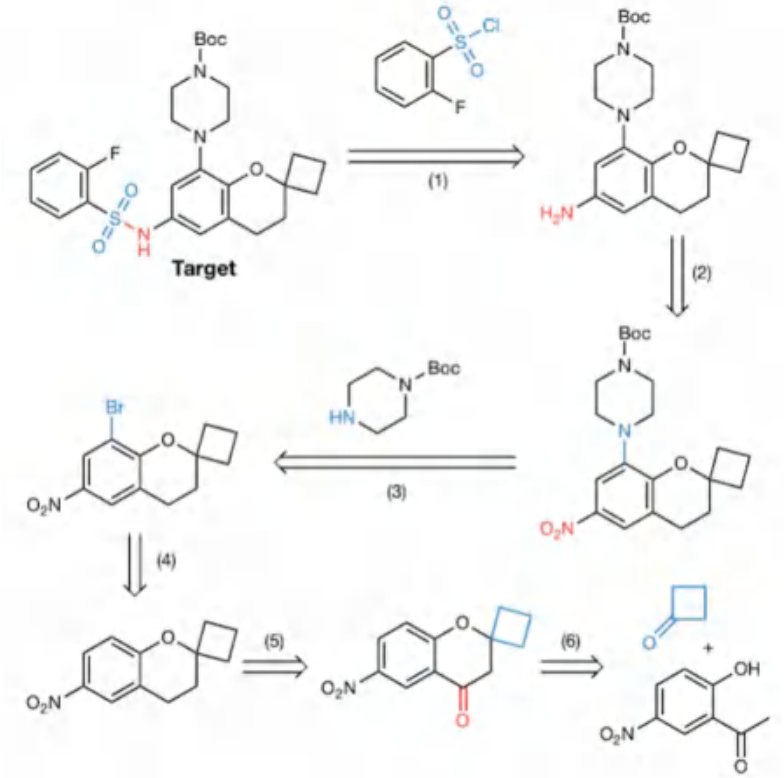
Gait analysis can identify early signs of dementia or Parkinson's



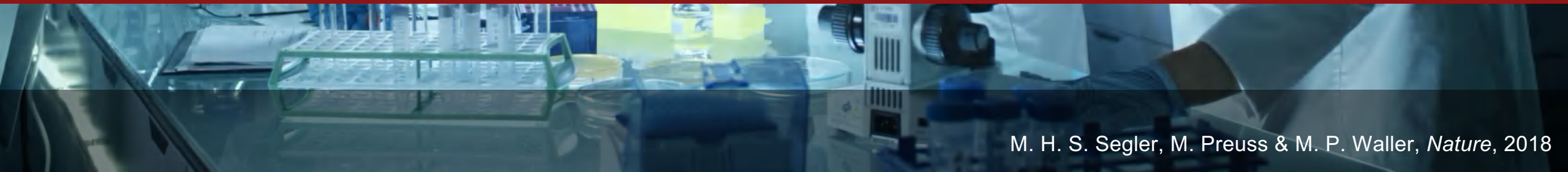
Yogev et al. Experimental Brain Research, 2007.


Lowers costs
Improves safety and outcomes
Gives caretaker time/effort back to patients

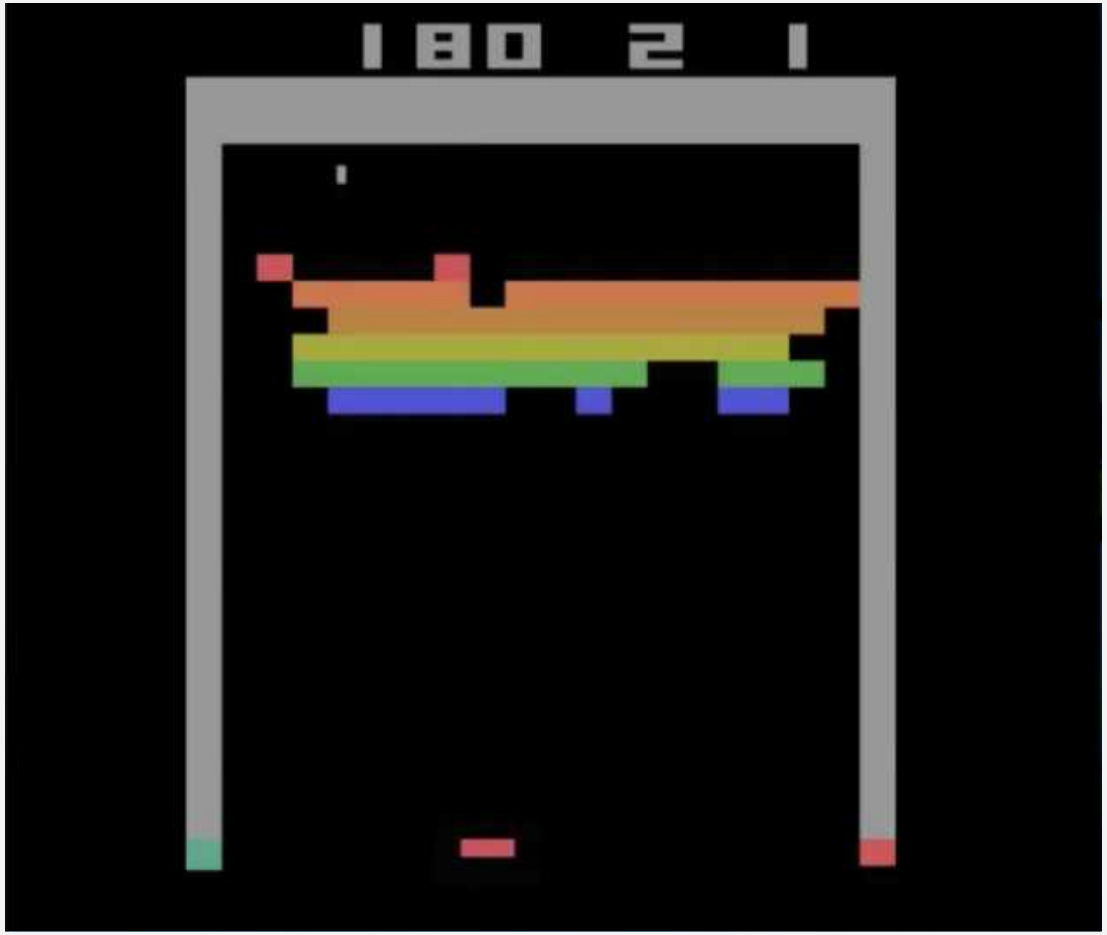

da Vinci S HD
SURGICAL SYSTEM

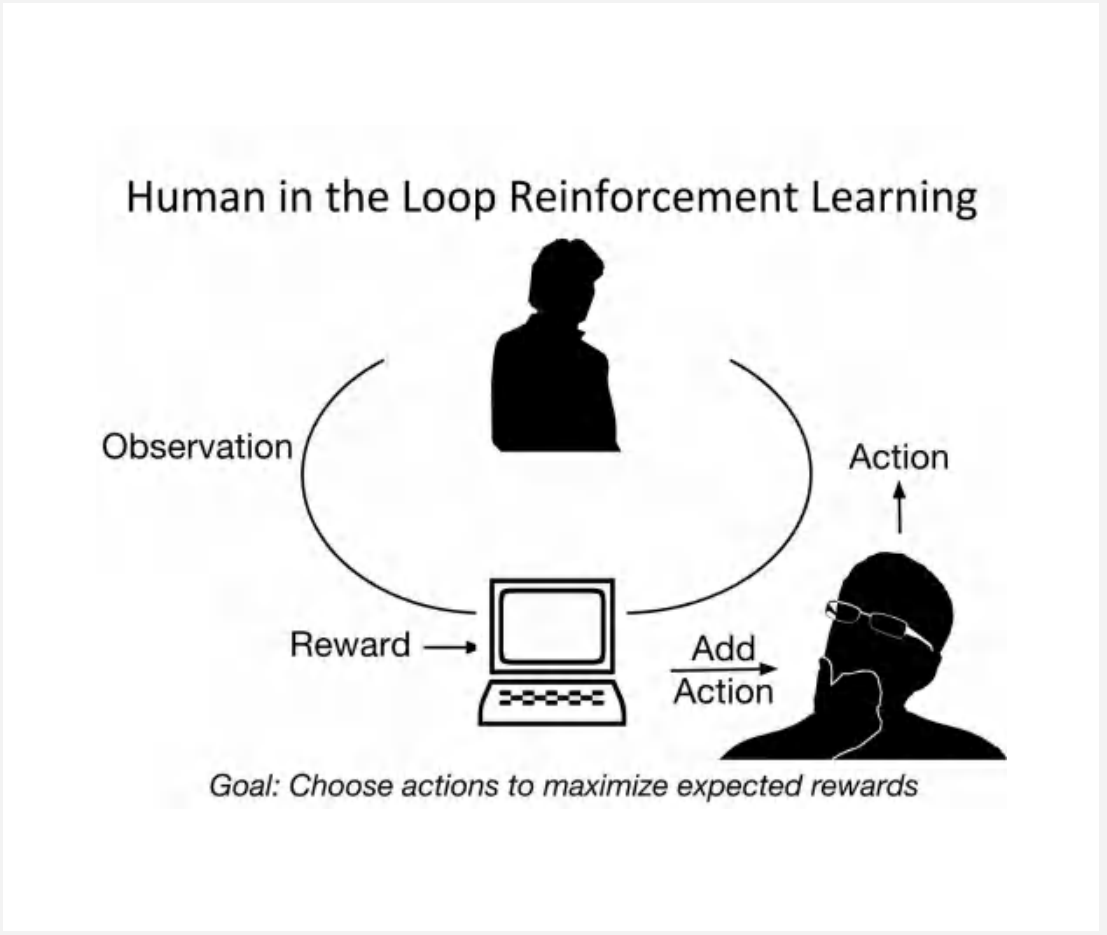


Algorithm-assisted research



M. H. S. Segler, M. Preuss & M. P. Waller, *Nature*, 2018

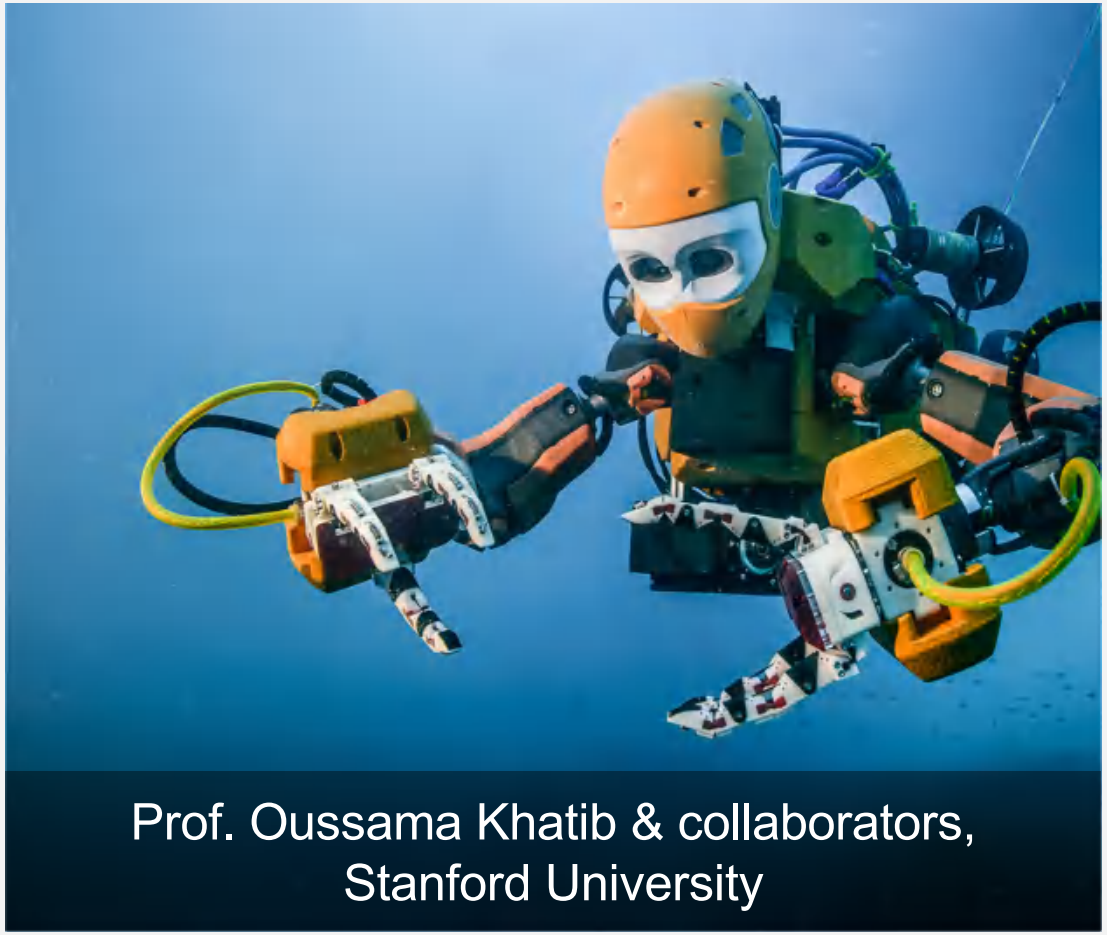




Reinforcement Learning in Education

Prof. Emma Brunskill & collaborators, Stanford University

Stanford HAI



Prof. Oussama Khatib & collaborators,
Stanford University

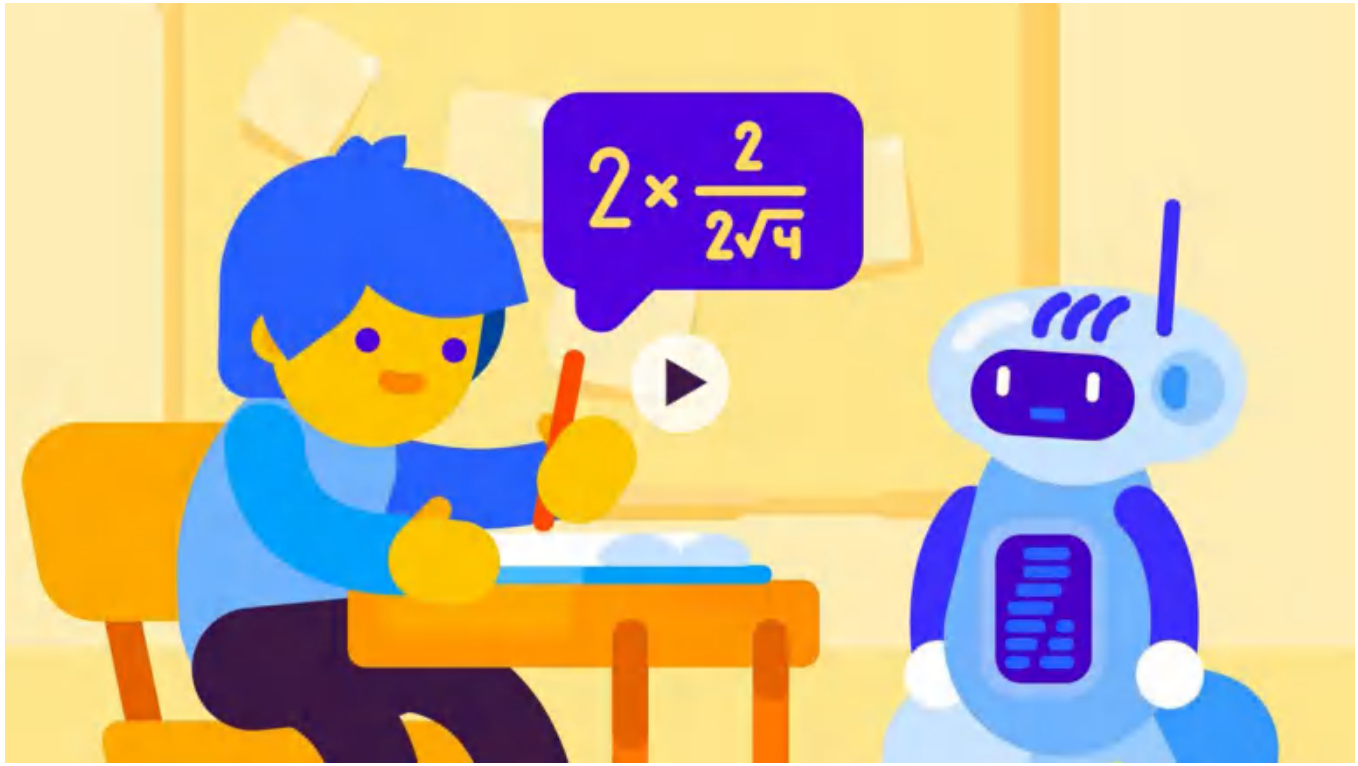
Today:
The Ocean One Robot can search the ocean at inhuman depths

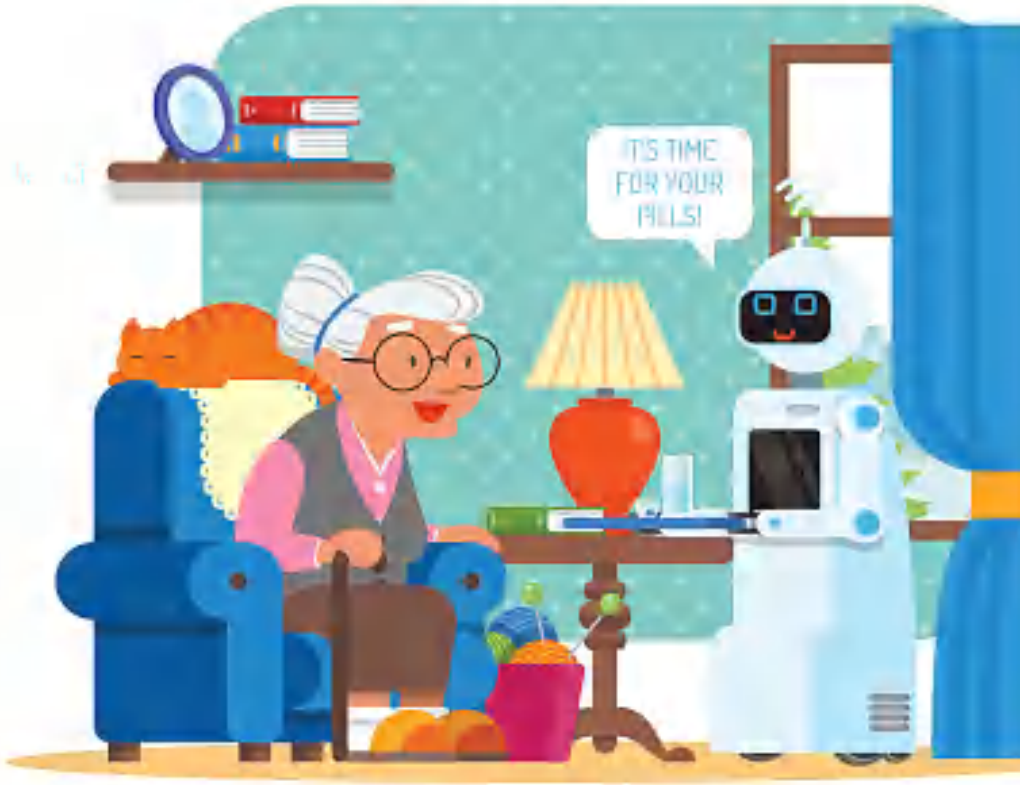
Tomorrow:
Intelligent machines will spare first responders in disaster areas

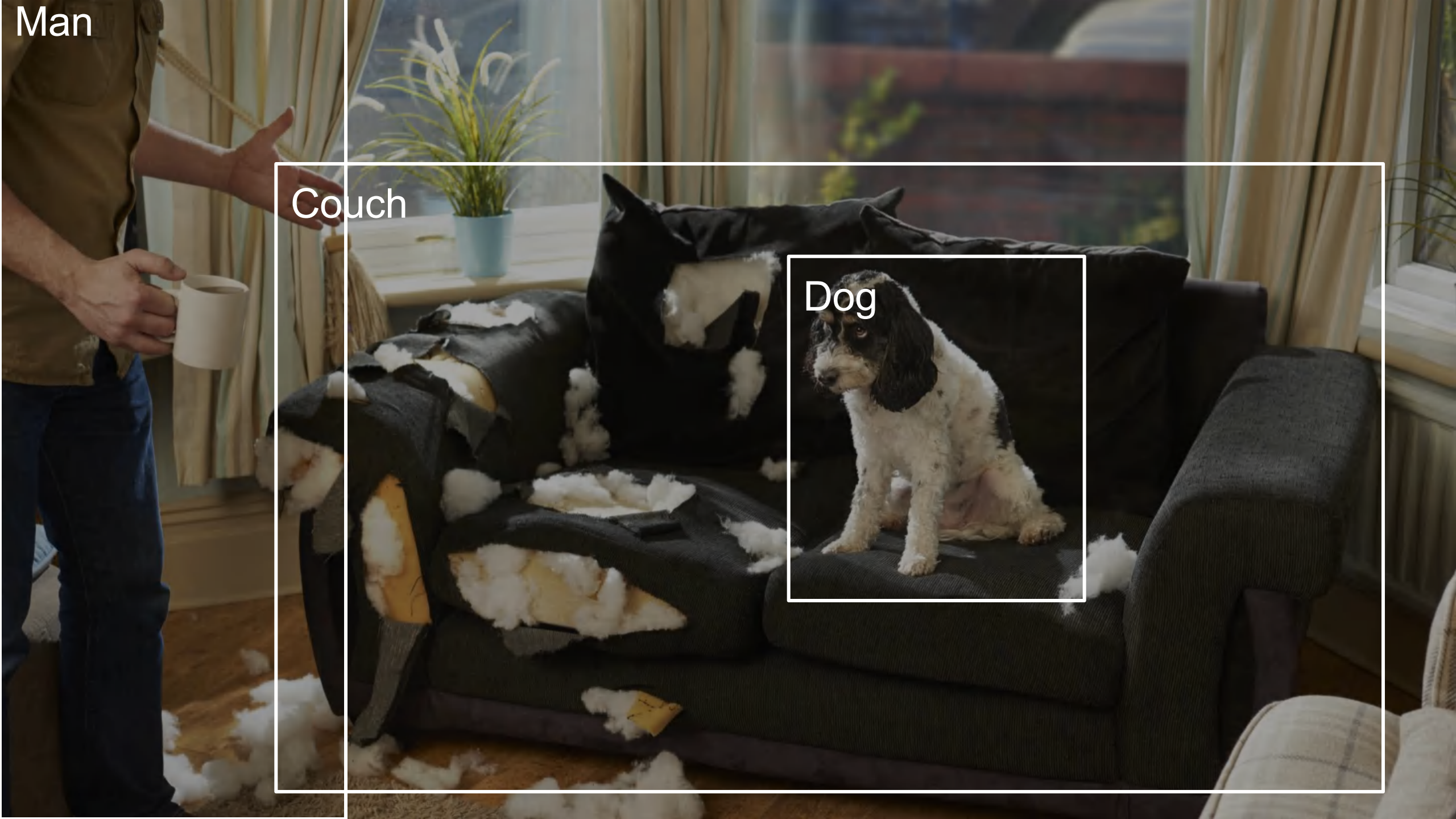
Stanford HAI

Human-Centered AI

AI must be more inspired by human **intelligence**.

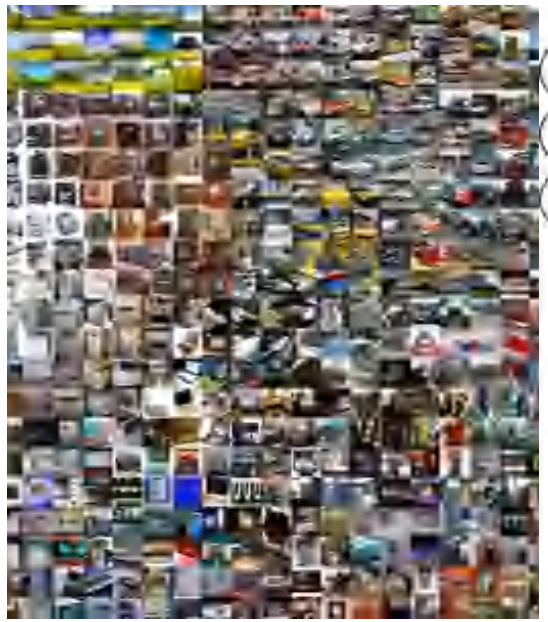

$2 \times \frac{2}{2\sqrt{4}}$

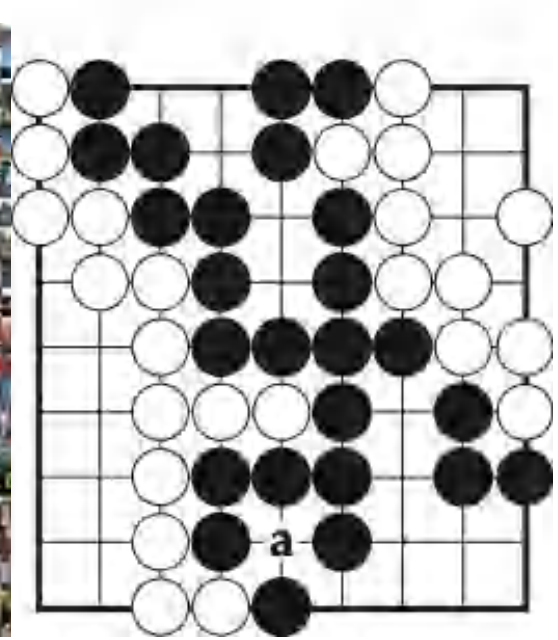

IT'S TIME FOR YOUR PILLS!


Man
Couch
Dog

Today's AI

Static, Simple goals,
Disembodied



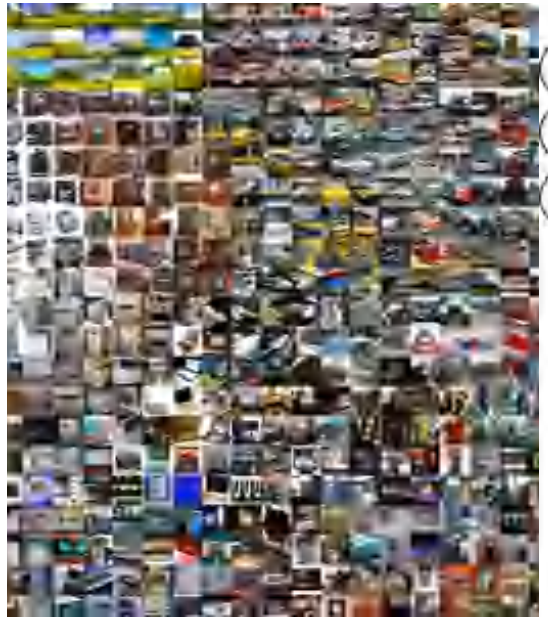


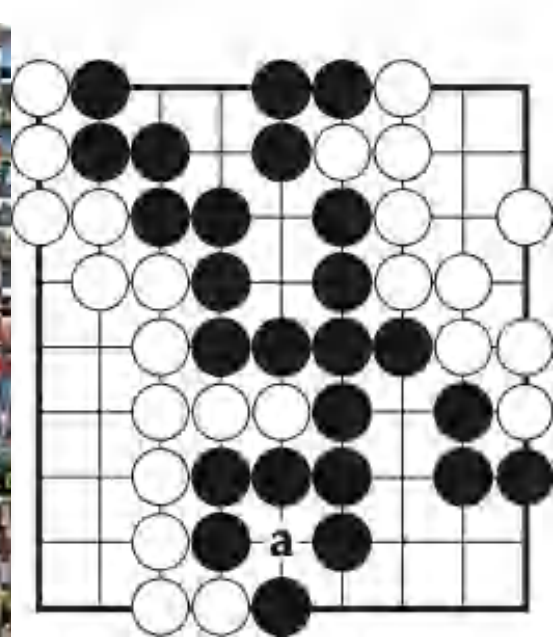
Human

Dynamic, Multi-sensory, Complex,
Uncertain, Interactive

Today's AI

Static, Simple goals,
Disembodied

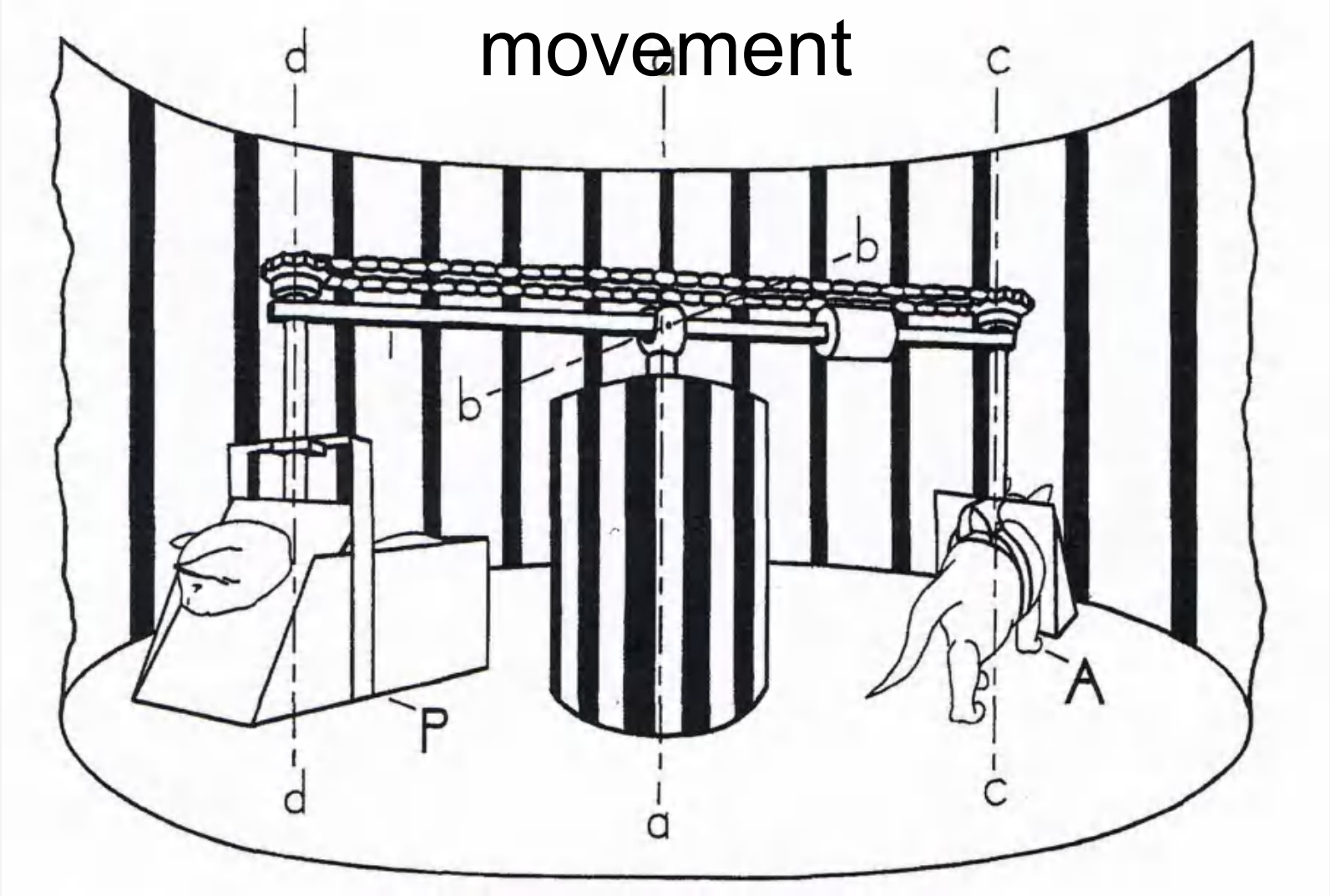




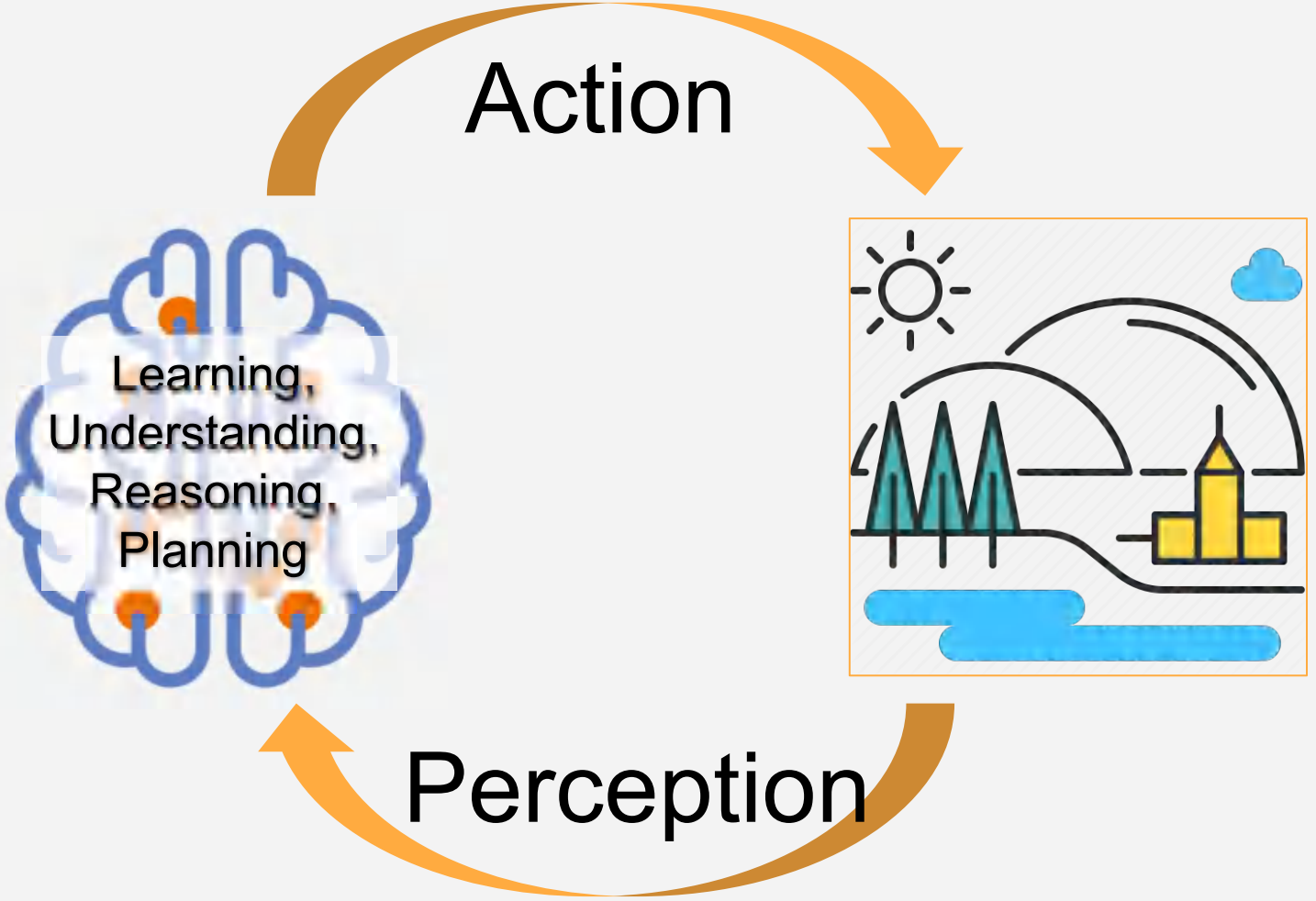
Human

Dynamic, Multi-sensory, Complex,
Uncertain, **Interactive**

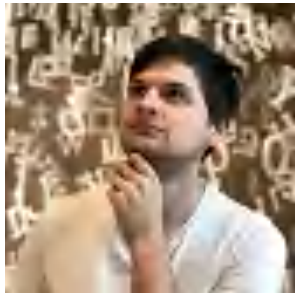
Held & Hein, 1963: Visual development requires self-guided movement



Intelligence emerges from **active perception** and **interaction** with the **real-world**



Interact like a **baby**



Damian Mrowca
PhD student

Prof. Nick Haber

Prof. Li Fei-Fei

Prof. Dan Yamins

Stanford HAI

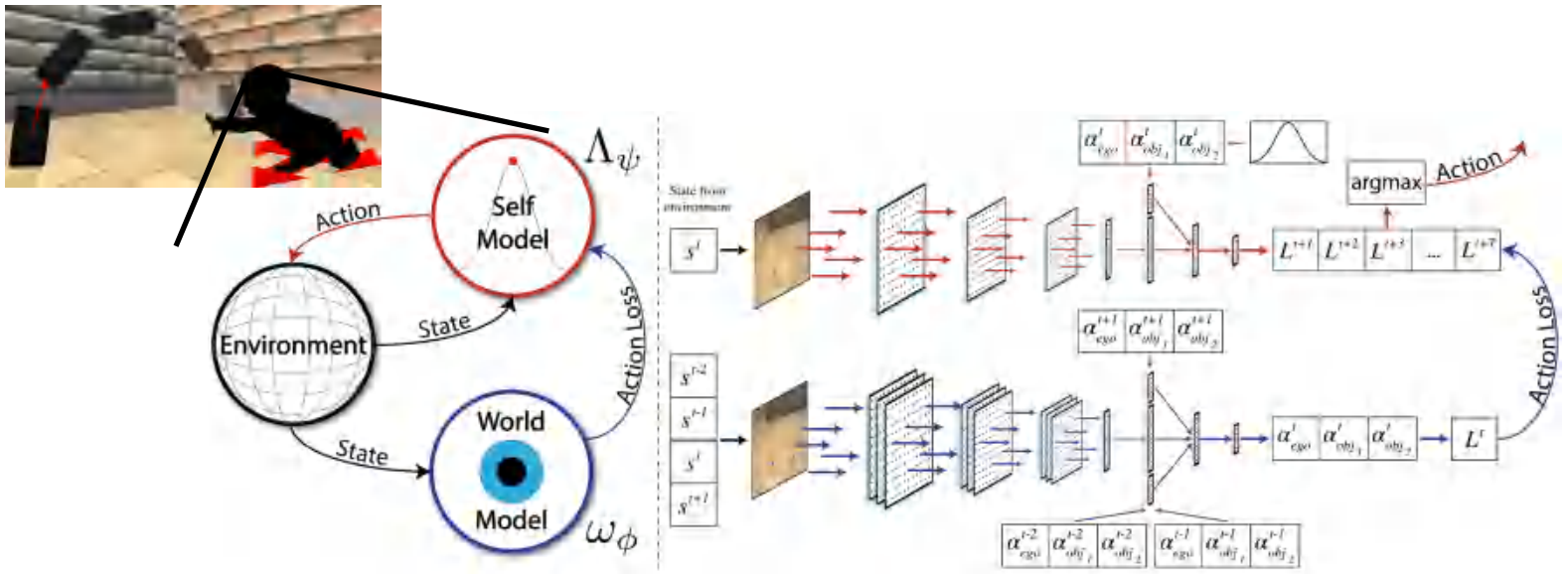
Infants are curious and play with their environment

How might an AI agent train to interact with its environment

- "Scientists in the crib!"
  Gopnik 2000
- Novelty preference
  Fantz 1964
- Goldilocks effect
  Kidd 2012

Haber & Mrowca et al. Cog, Sci. 2018

Dynamics learning through intrinsically motivated interactions



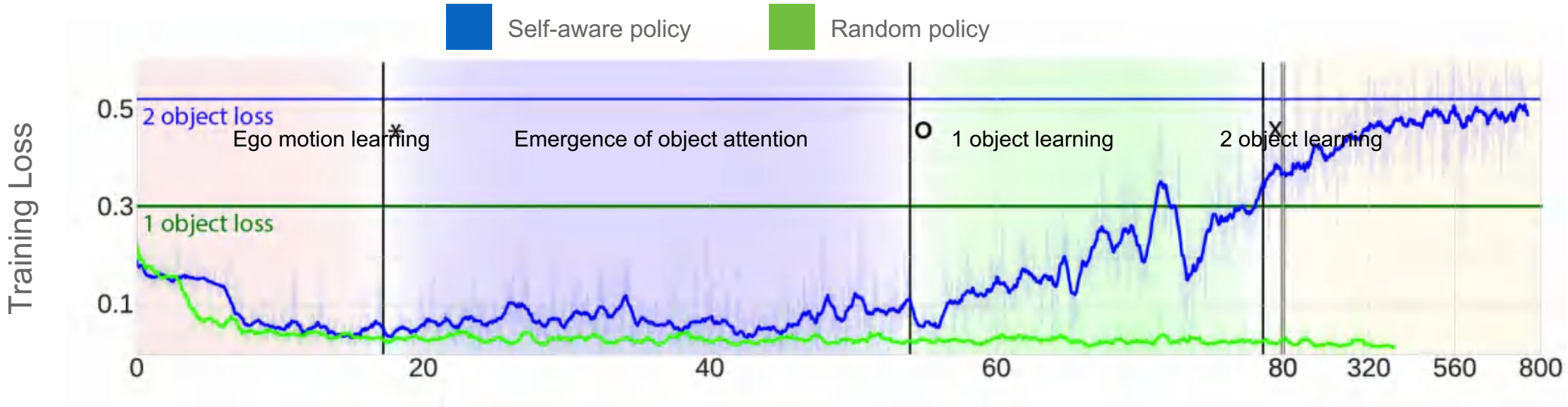
World Model network:
to predict consequences of actions

Self Model network:
to predict errors of world-model (“self-aware”)

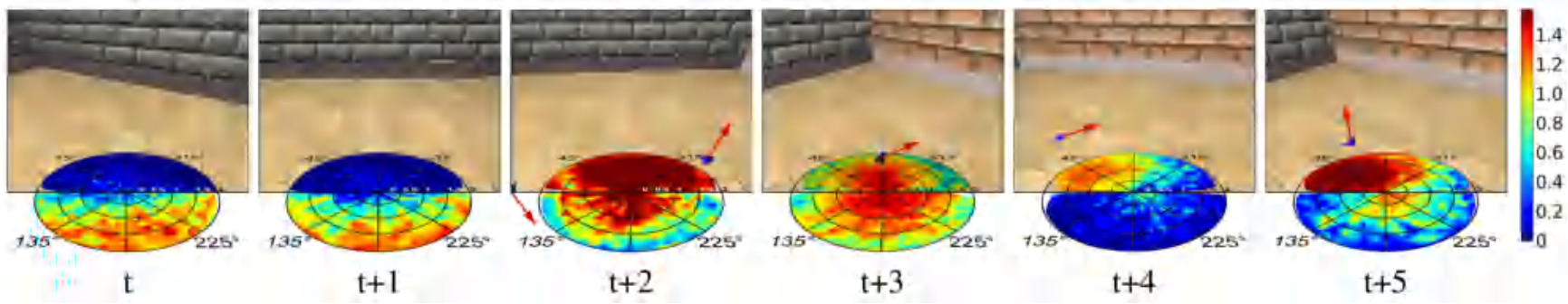
Action choice:
self-model is **adversarial** to world-model (“curious intrinsic motivation”)

Haber & Mrowca et al. NeurIPS, 2018

Emergent Behavior: **Agent goes through learning stages**

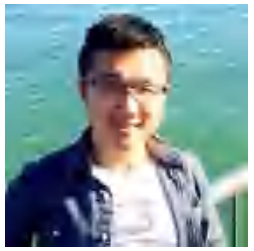


Primitive planning: **Agent learns to focus on objects**



Haber & Mrowca et al. NeurIPS, 2018

Interact with the **physical environment**



Kuan Fang
PhD student

Prof. Yuke Zhu
Former PhD student

Prof. Animesh Garg
Former PostDoc

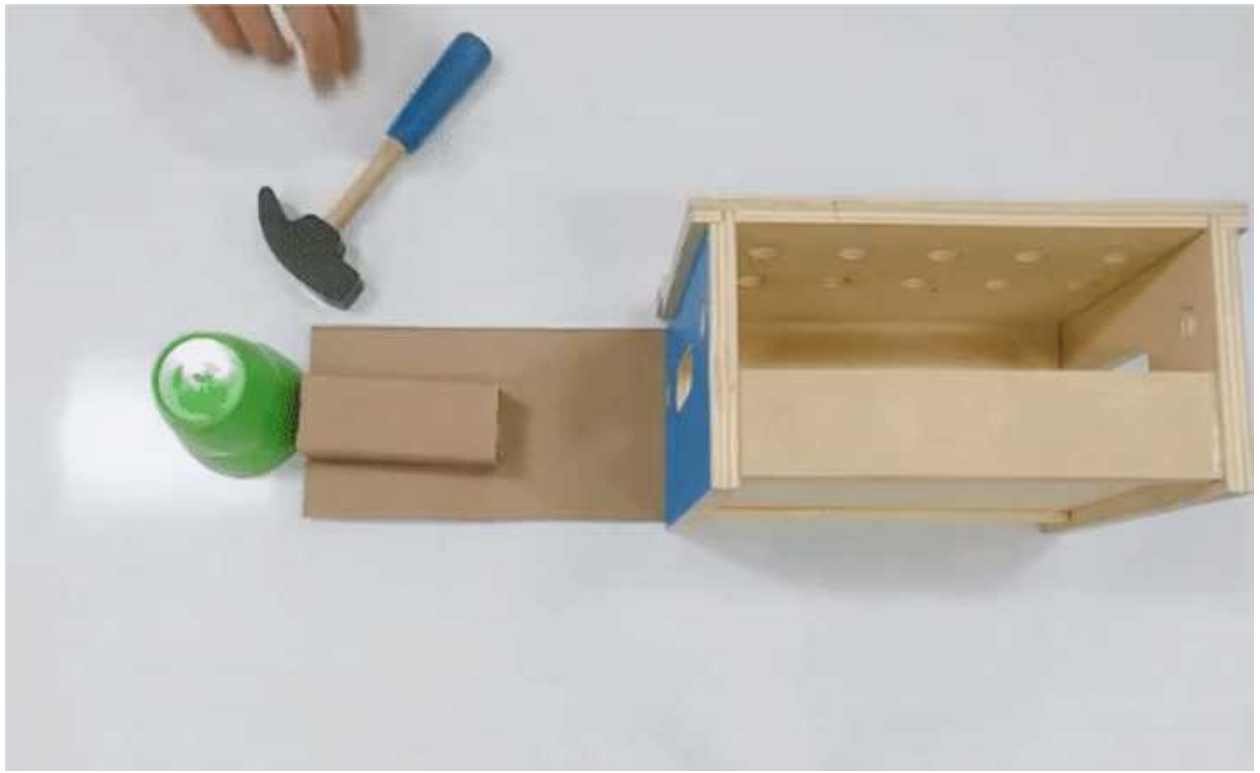
Prof. Li Fei-Fei

Prof. Silvio Savarese

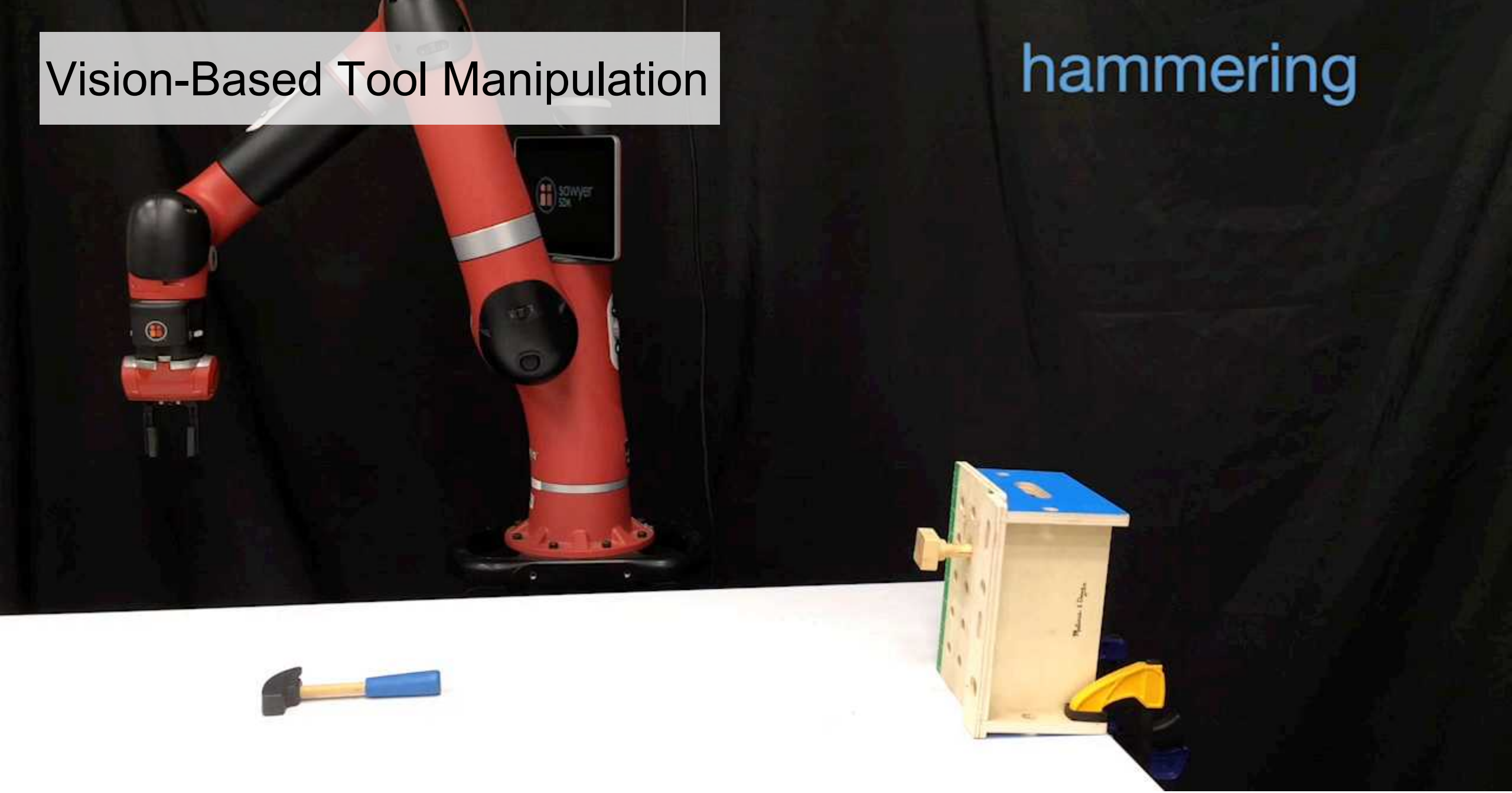
Stanford HAI

Vision-Based Tool Manipulation

Recognition > Understanding > Manipulation

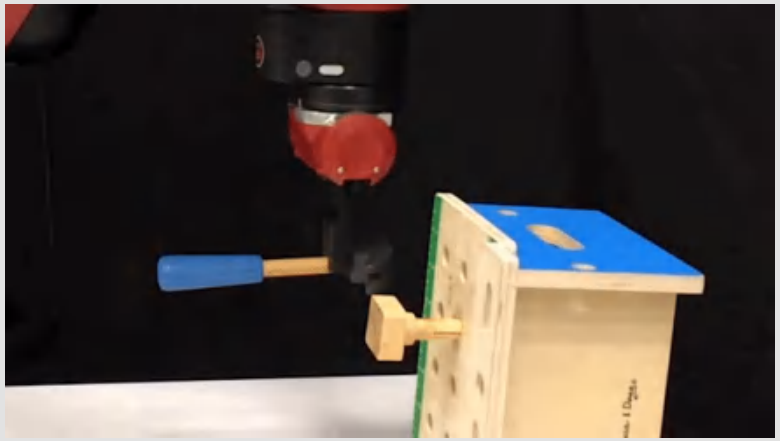
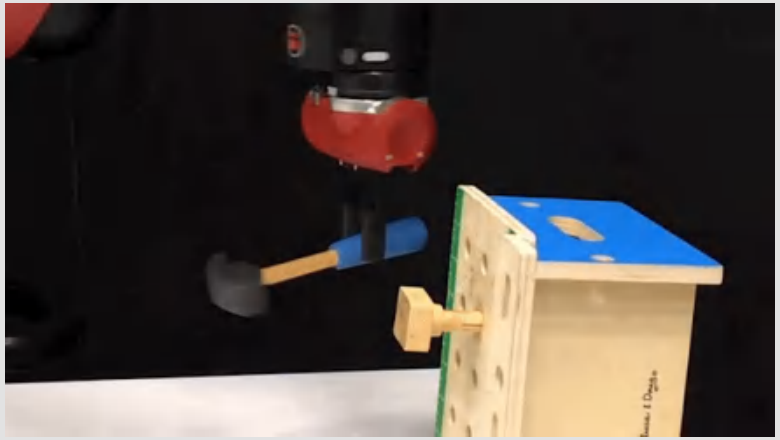


Vision-Based Tool Manipulation



K. Fang, Y. Zhu, A. Garg, A. Kurenkov, V. Mehta, L. Fei-Fei, S. Savarese. RSS'18

Task-Agnostic Grasp vs. Task-Oriented Grasp

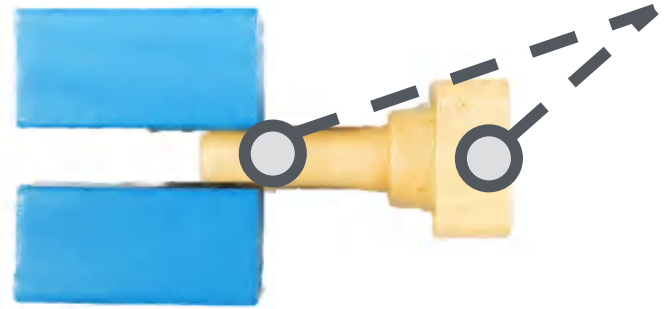
Task-Agnostic Grasp		
Task-Oriented Grasp		

K. Fang, Y. Zhu, A. Garg, A. Kurenkov, V. Mehta, L. Fei-Fei, S. Savarese. RSS'18

Keypoint Representations for Tool Manipulation

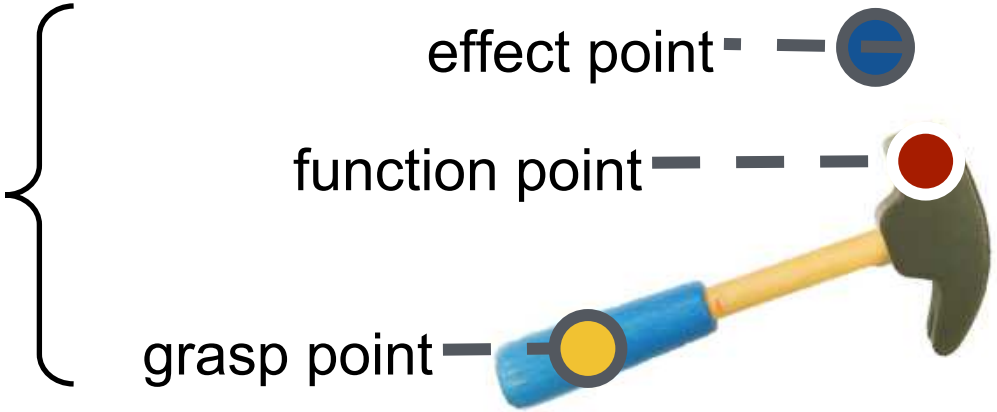
Fixed

environment keypoints



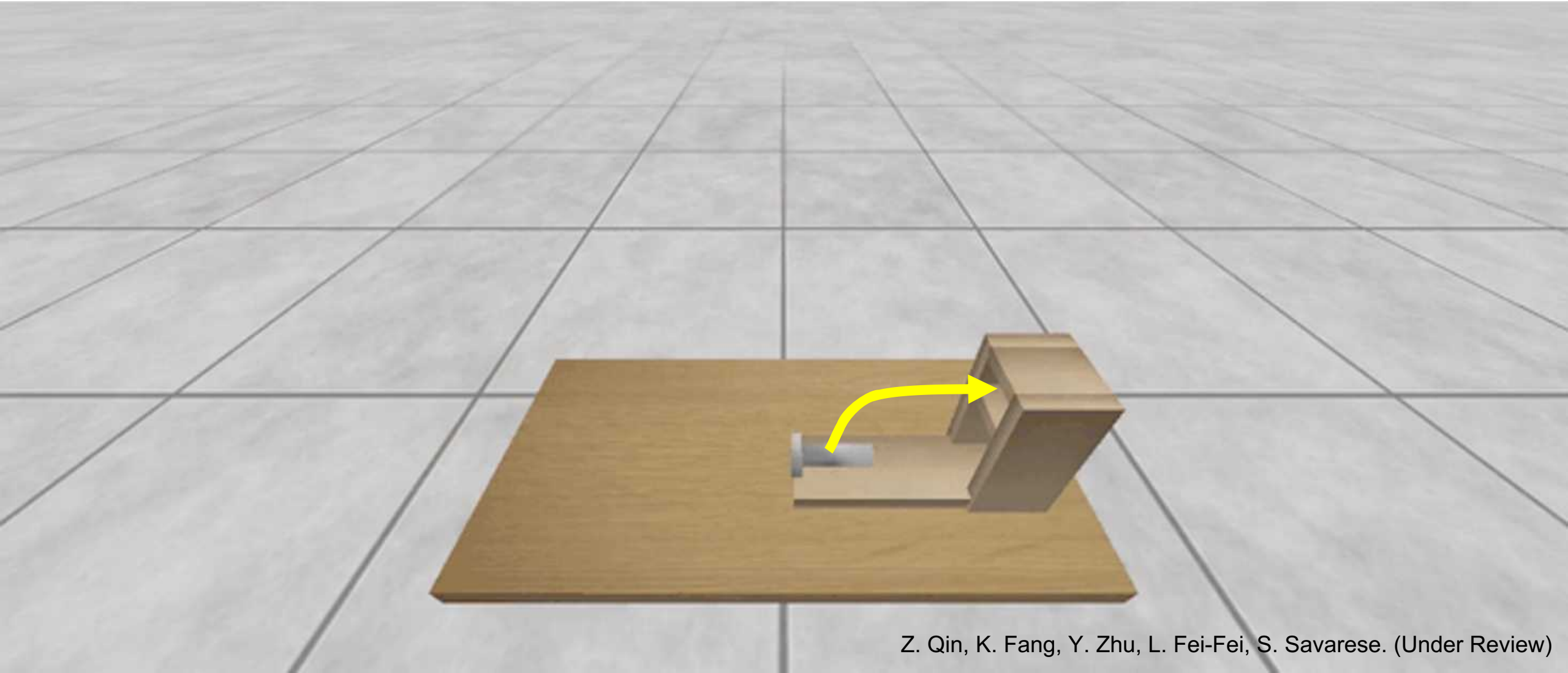
Learned

tool keypoints



Z. Qin, K. Fang, Y. Zhu, L. Fei-Fei, S. Savarese. (Under Review)

Composite Task: Multi-Stage Tool Use



Z. Qin, K. Fang, Y. Zhu, L. Fei-Fei, S. Savarese. (Under Review)

Tool Creation

Keypoints offer a template for generating tools from object parts.

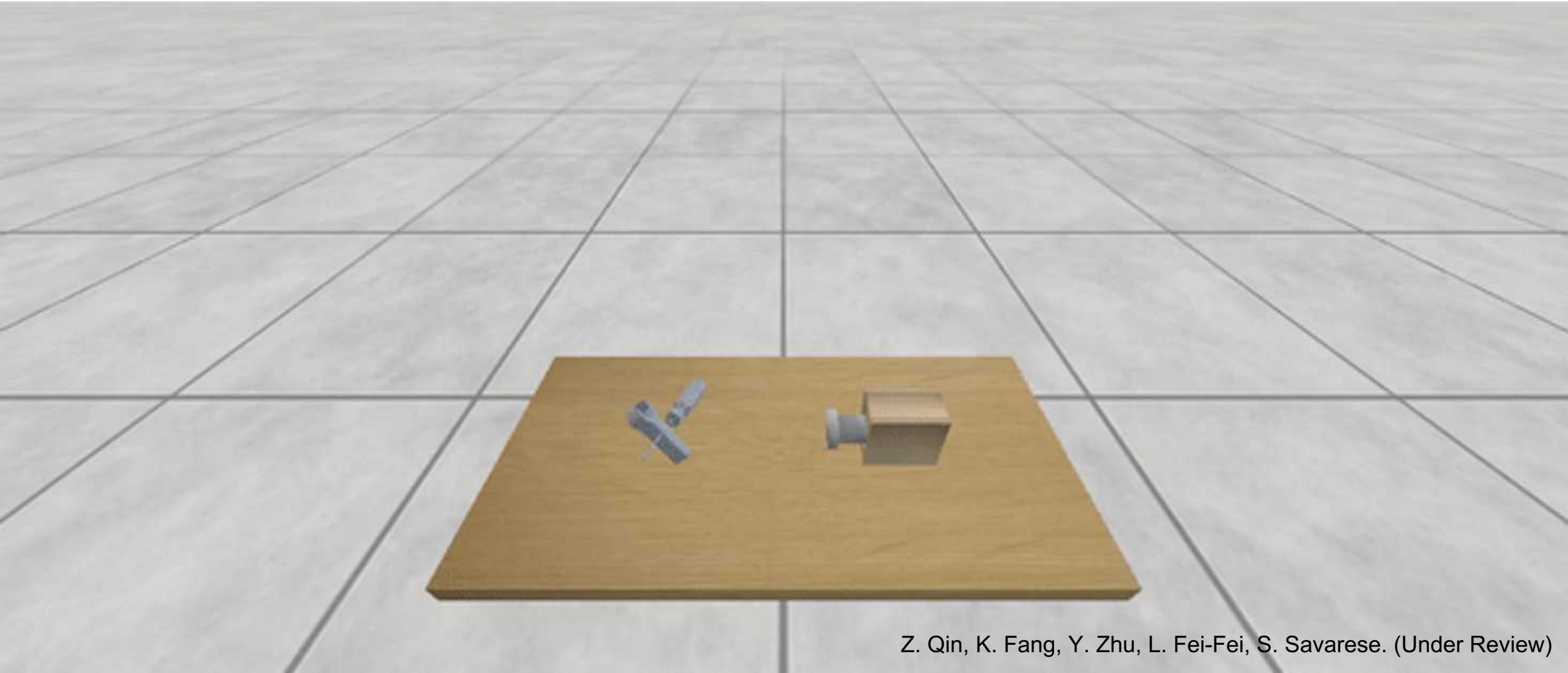
Pushing

Reaching

Hammering

Z. Qin, K. Fang, Y. Zhu, L. Fei-Fei, S. Savarese. (Under Review)

Hammering with the Created Tool

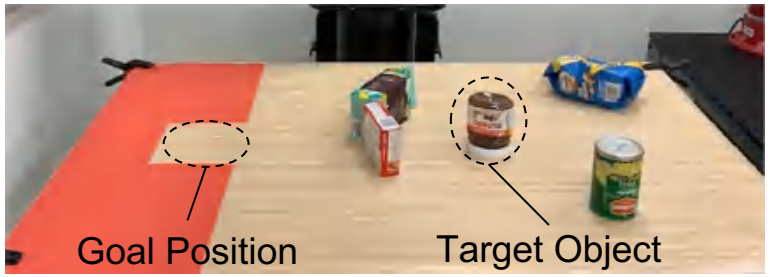


Z. Qin, K. Fang, Y. Zhu, L. Fei-Fei, S. Savarese. (Under Review)

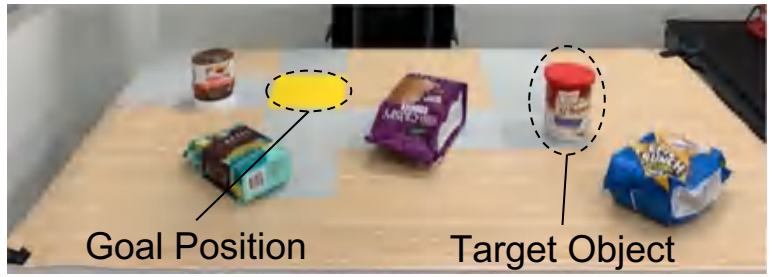
Generalizable Multi-Step Manipulation for Various Tasks and Targets

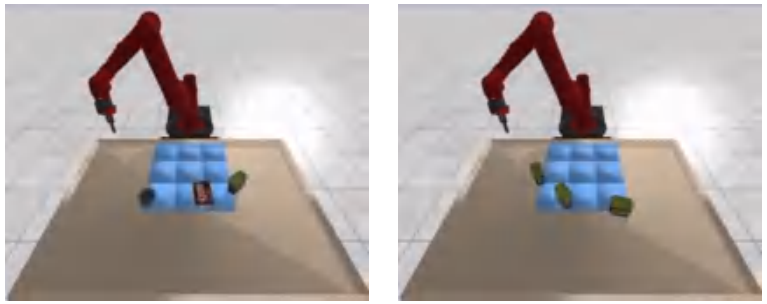
Clearing

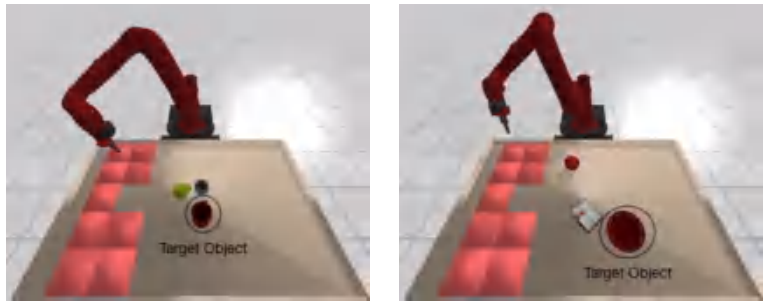
Insertion

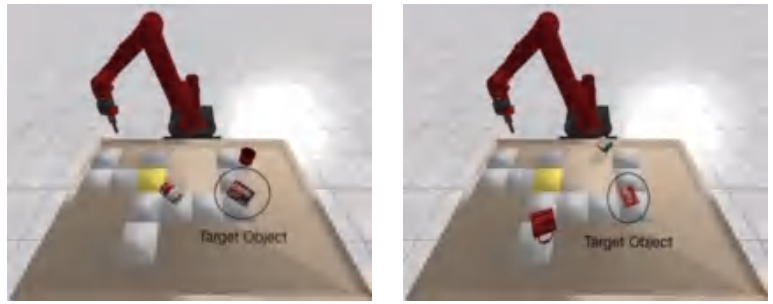


Crossing



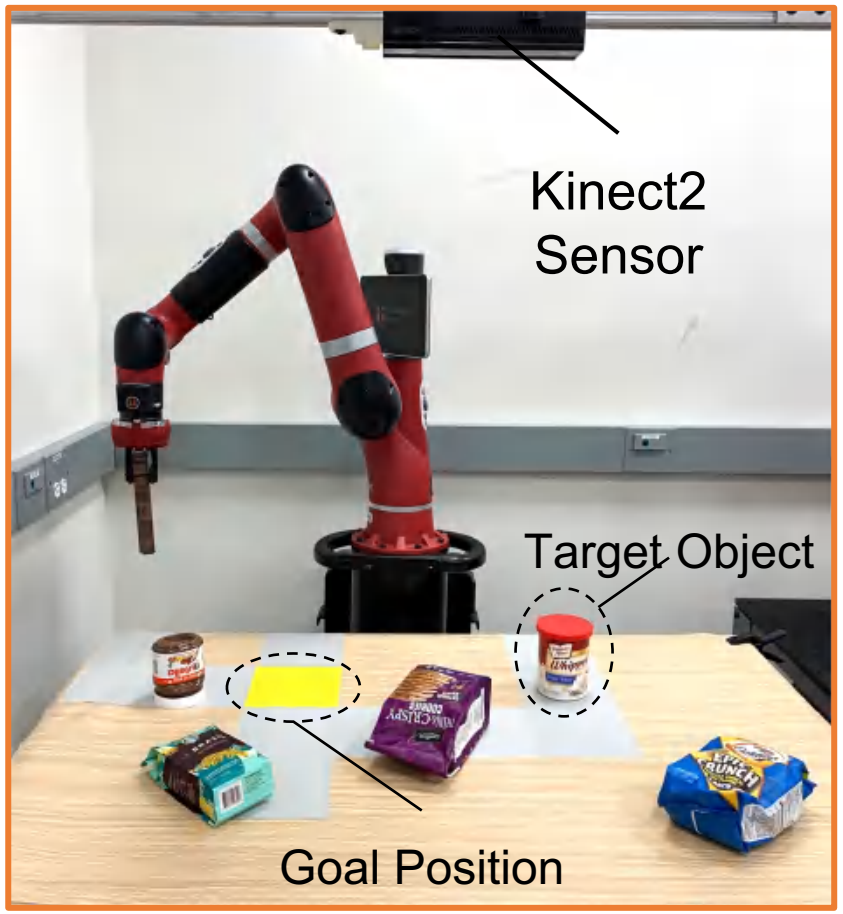


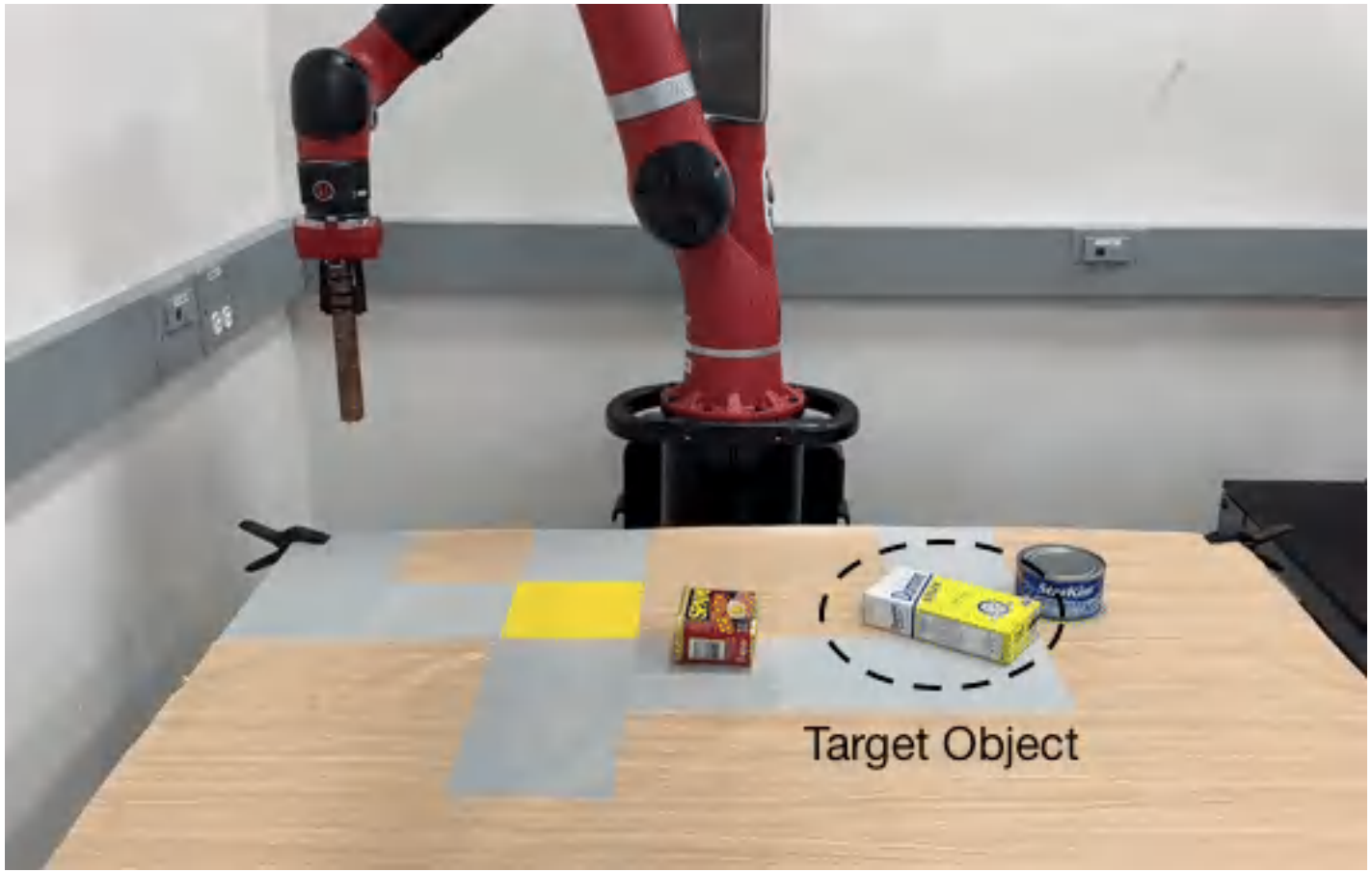




K. Fang, Y. Zhu, A. Garg, S. Savarese, L. Fei-Fei. CoRL'19

Generalizable Multi-Step Manipulation for Various Tasks and Targets





K. Fang, Y. Zhu, A. Garg, S. Savarese, L. Fei-Fei. CoRL'19

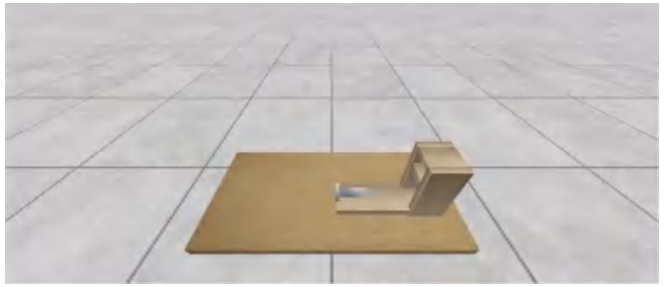
Reasoning about Tools for Manipulation

Task-oriented grasping for tool manipulation

object | task-agnostic | sweeping | hammering

Fang, Zhu, Garg, Kurenkov, Fei-Fei & Savarese., RSS 2018

Multi-stage tool use and creation



Qin, Fang, Zhu, Fei-Fei & Savarese (under review)

Observational Learning for Task Structures

Exploiting hierarchical task structures for better generalization

video demonstration | one-shot imitation learning | real robot execution

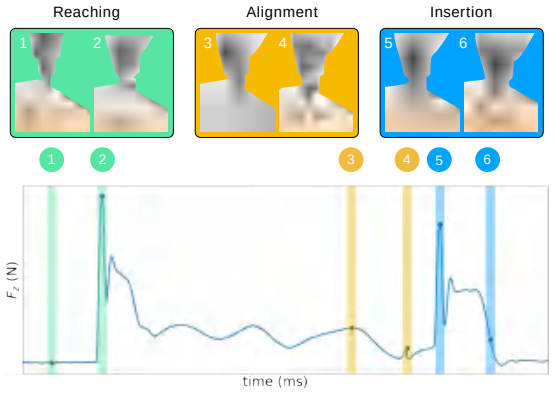
Xu*, Nair*, et al., ICRA 2018; Huang*, Nair*, Xu*, et al., CVPR 2019

Multimodal Learning with Vision and Force

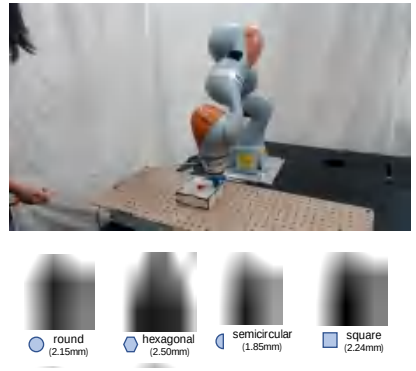
Learning multimodal representations for high-precision manipulation

vision

force



robust policy



Lee*, Zhu*, et al., ICRA 2018 (Best Paper Award)

Interact with **humans**

Ranjay Krishna
PhD student

Prof. Michael Bernstein

Prof. Li Fei-Fei

Stanford HAI

Humans learn by interacting with other humans

Our aim:
A conversational AI agent that **learns visual knowledge** by **interacting** with and learning from people

Sarah: Found this little one outside of my apartment.

Red Panda.

Sarah: Found this little one outside of my apartment.

Agent: I have never seen that animal, is that a fox?

Sarah: Found this little one outside of my apartment.

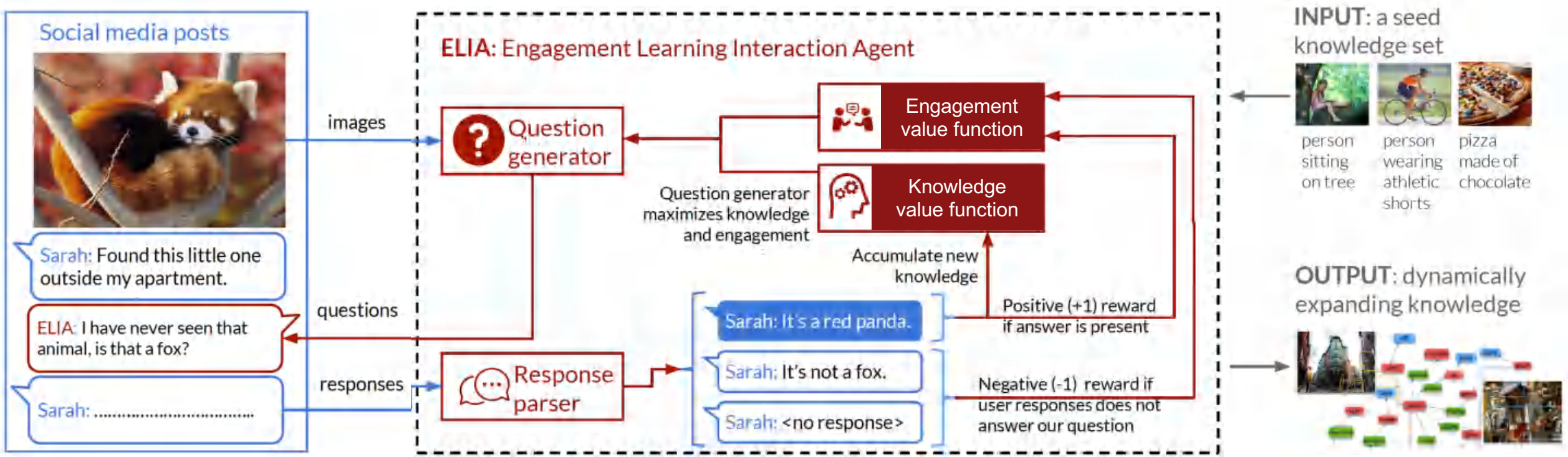
Agent: I have never seen that animal, is that a fox?

Sarah: It's a **Red Panda.**

Engagement Learning:
an open-vocabulary reinforcement learning algorithm

Krishna et al. submitted

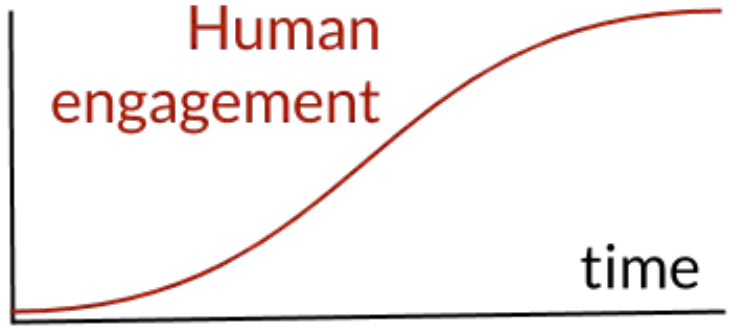
Engagement Learning:
an open-vocabulary reinforcement learning algorithm



Krishna et al. submitted

Engagement Learning: Results

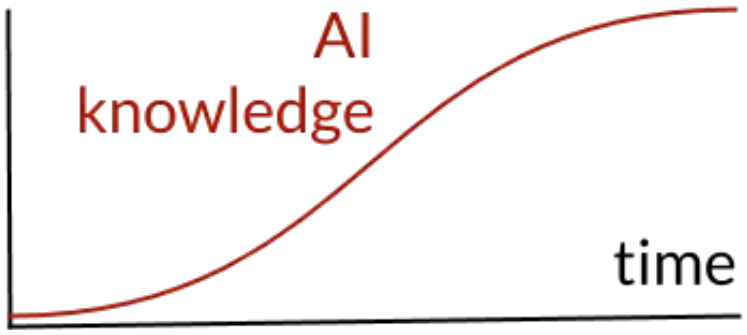
Goal 1:
Ask More Engaging Questions



Krishna et al. submitted

Engagement Learning: Results

Goal 2:
Expand visual knowledge
of the AI agent



Q: What kind of bird is that?

A (by AI): Magpie

Q: What kind of flower is that?

A (by AI): Dahlias

Q: What is the white stuff on the plate?

A (by AI): Feta cheese

Krishna et al. submitted

Human-Centered AI

H

The development of AI must be guided by a concern for its **human impact**.

A

AI should strive to **augment** and enhance us, not replace us.

I

AI must be more inspired by human **intelligence**.

Human-Centered AI

Thank you!