The Rise of AI
And The Challenges of Human-Aware AI Systems

Subbarao Kambhampati

Arizona State University

rao@asu.edu @rao2z @subbarao2z

WeChat: Subbarao2z

CCF-GAIR, Shenzen, July 7th, 2017

AAAI & China AI Community

- Founded in 1979, AAAI is the oldest and largest scientific society devoted to AI
- Researchers from China are a formidable force in AAAI
  - Rivals USA in terms of paper submission and acceptance
  - AAAI-17 dates shifted to avoid conflict with the start of the Year of Rooster!
  - Prof. Qiang Yang is on the Executive Council
  - Prof. Zhi-Hua Zhou is co-program-chair for AAAI 2019
- AAAI welcomes even more vigorous participation from China AI community
  - Only one in 23 AAAI members are from China (USA: 1 in 2; UK: 1 in).
  - 20$/year membership for China.
  - Join AAAI!

China's Artificial-Intelligence Boom

The country's universities and tech giants are starting to surpass American ones when it comes to researching and implementing AI.

SARAH ZHANG | FEB 16, 2017 | TECHNOLOGY

Share Tweet ...

TEXT SIZE − +

32nd AAAI Conference in February 2-7, 2018 in New Orleans!

Increasing International Representation

AAAI-16 Registrants by Country

Row Labels	Count of Country
Argentina	1
Australia	33
Austria	7
Belgium	6
Brazil	4
Bulgaria	1
Canada	50
China	156
Czech Republic	5
Denmark	2
Finland	6
France	13
Germany	18
Greece	2
Hong Kong	22
India	21
Ireland	5
Israel	19
Italy	13
Japan	48
Kazakhstan	2
Korea, Republic of	15
Netherlands	5
Norway	2
Pakistan	1
Poland	2
Portugal	1
Romania	2
Russian Federation	1
Saudi Arabia	5
Singapore	34
South Africa	1
Spain	2
Sweden	5
Switzerland	15
Taiwan	6
Turkey	4
United Arab Emirates	2
United Kingdom	44
United States	708
Grand Total	**1289**

Country	2016	2017
USA	708	964 (up 27%)
China	156	275 (up 76%)
Japan	48	92
Korea	15	68
UK	44	51

AAAI-17 Registrants by Country

Row Labels	Count of Country
Argentina	1
Australia	34
Austria	9
Belgium	9
Brazil	6
Canada	47
Chile	1
China	275
Czech Republic	5
Denmark	2
Ecuador	1
Finland	1
France	20
Germany	37
Greece	2
Hong Kong	17
Iceland	1
India	31
Ireland	5
Israel	22
Italy	12
Japan	92
Korea, Republic of	68
Netherlands	9
New Zealand	1
Norway	2
Pakistan	1
Philippines	1
Poland	6
Portugal	2
Romania	1
Russian Federation	3
Saudi Arabia	2
Singapore	42
South Africa	3
Spain	2
Sweden	7
Switzerland	11
Taiwan	10
United Kingdom	51
United States	964
Venezuela	2
(blank)	
Grand Total	**1818**

AAAI Chapters

- There has been some resurgent interest in AAAI chapters
  - Given the significant public interest in AI, the AAAI council is very supportive of chapters
  - (Distinguished Speaker Program)

AAAI Chapter Program

AAAI is delighted to announce the establishment of the AAAI Chapter Program. AAAI chapters will be organized and operated for charitable, educational, and scientific purposes to promote the nonprofit mission of AAAI, including:

a. Increasing knowledge of, and greater interest in, artificial intelligence (AI);
b. Promoting greater awareness of AI and its potential among the chapter's local community at large;
c. Fostering greater interactions between the chapter AI community and the international AI community; and
d. Promoting greater participation of and membership in AAAI by chapter members.

Complete AAAI Chapter Guidelines are available below. If you are interested in starting a AAAI Chapter, please contact chapters13@aaai.org, or fill out the application form below.

AAAI Chapter Application Form
(https://www.aaai.org/Forms/chapter-form.php)

AAAI & China AI Community

- Founded in 1979, AAAI is the oldest and largest scientific society devoted to AI
- Researchers from China are a formidable force in AAAI
  - Rivals USA in terms of paper submission and acceptance
  - AAAI-17 dates shifted to avoid conflict with the start of the Year of Rooster!
  - Prof. Qiang Yang is on the Executive Council
  - Prof. Zhi-Hua Zhou is co-program-chair for AAAI 2019
- AAAI welcomes even more vigorous participation from China AI community
  - Only one in 23 AAAI members are from China (USA: 1 in 2; UK: 1 in).
  - 20$/year membership for China.
  - Join AAAI!

China's Artificial-Intelligence Boom

The country's universities and tech giants are starting to surpass American ones when it comes to researching and implementing AI.

SARAH ZHANG | FEB 16, 2017 | TECHNOLOGY

Share Tweet ...

TEXT SIZE − +

32nd AAAI Conference in February 2-7, 2018 in New Orleans!

SOME EXPERIMENTS
ON
ISOLATED WORD SPEECH RECOGNITION
FOR
CONFUSABLE VOCABULARY

A PROJECT REPORT
Submitted in partial fulfilment of the requirements for
the award of the degree of
BACHELOR OF TECHNOLOGY
in
ELECTRICAL ENGINEERING
(ELECTRONICS)

by

KAMBHAMPATI SUBBA RAO

Under the guidance of
Prof. B. YEGNANARAYANA

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
MADRAS - 600 036. INDIA.

I. INTRODUCTION :-

I.1 OBJECTIVE :-

The objective of this study is to investigate the performance of existing ISOLATED WORD RECOGNITION SYSTEMS for confusable vocabulary and to suggest methods for improving the performance.

I.2 EXISTING SYSTEMS:

Speech Recognition, as a very important problem of Pattern-Recognition has been recognised long back and efforts to make Speech Recognition a practical reality date as far back as 1950's (1). One of the very first problems, to be tackled in Speech Recognition is "Recognition of Isolated Words". Apart from being the simplest facet of Speech Recognition, IWR has been found to have potential commercial applications (2) and more importantly to be a first step towards more complicated problems of Connected Word Recognition and finally Speech Understanding.

I.2.1 DESCRIPTION OF EXISTING SYSTEMS:

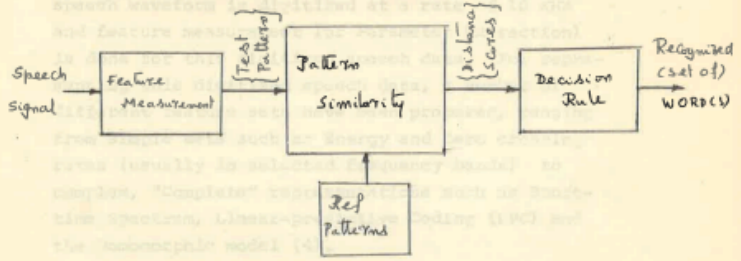
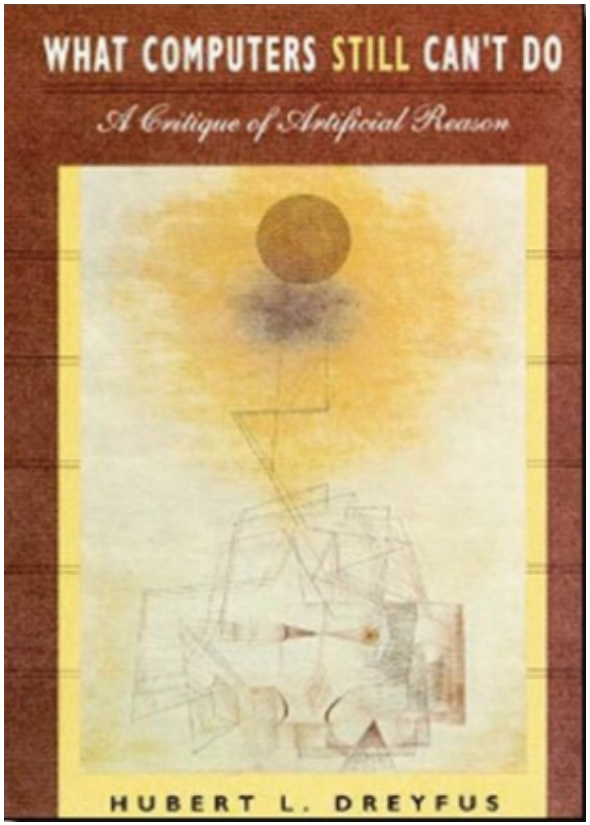
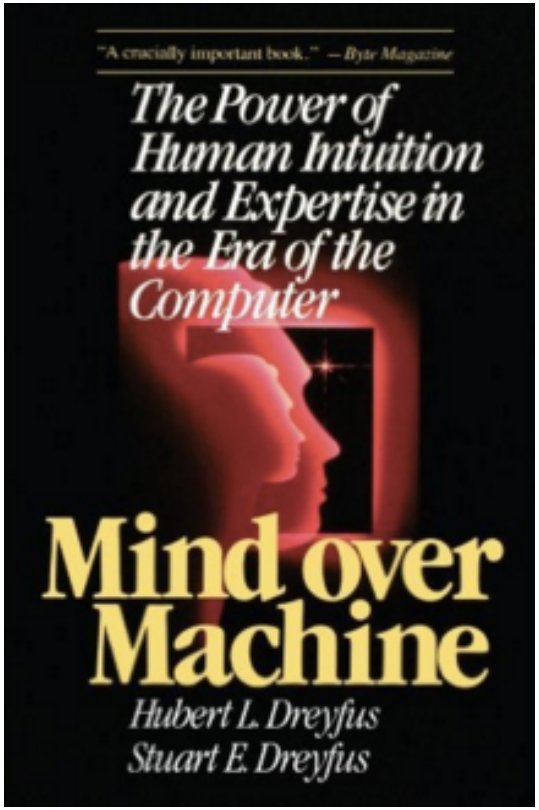


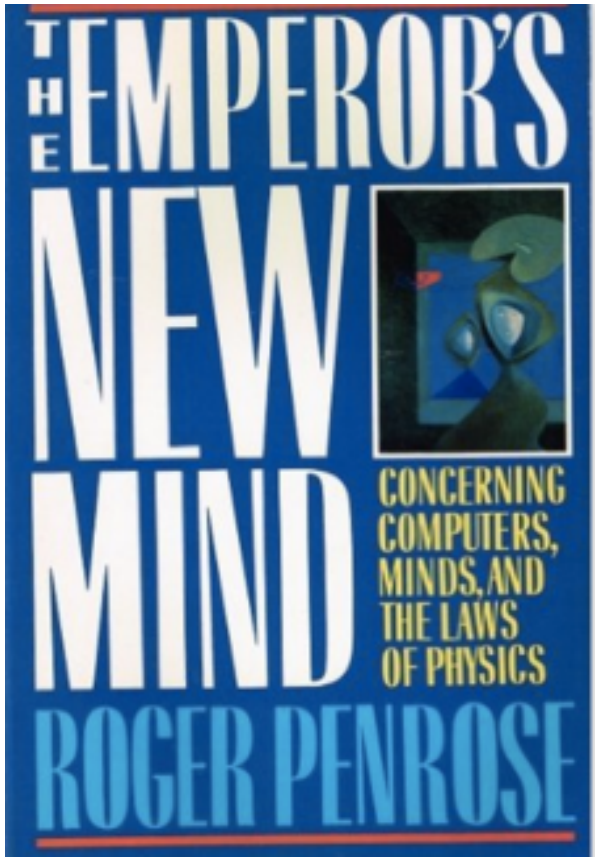
FIG.I.2.1 PATTERN-RECOGNITION MODEL FOR IWR:

1983 Bachelors thesis ☺

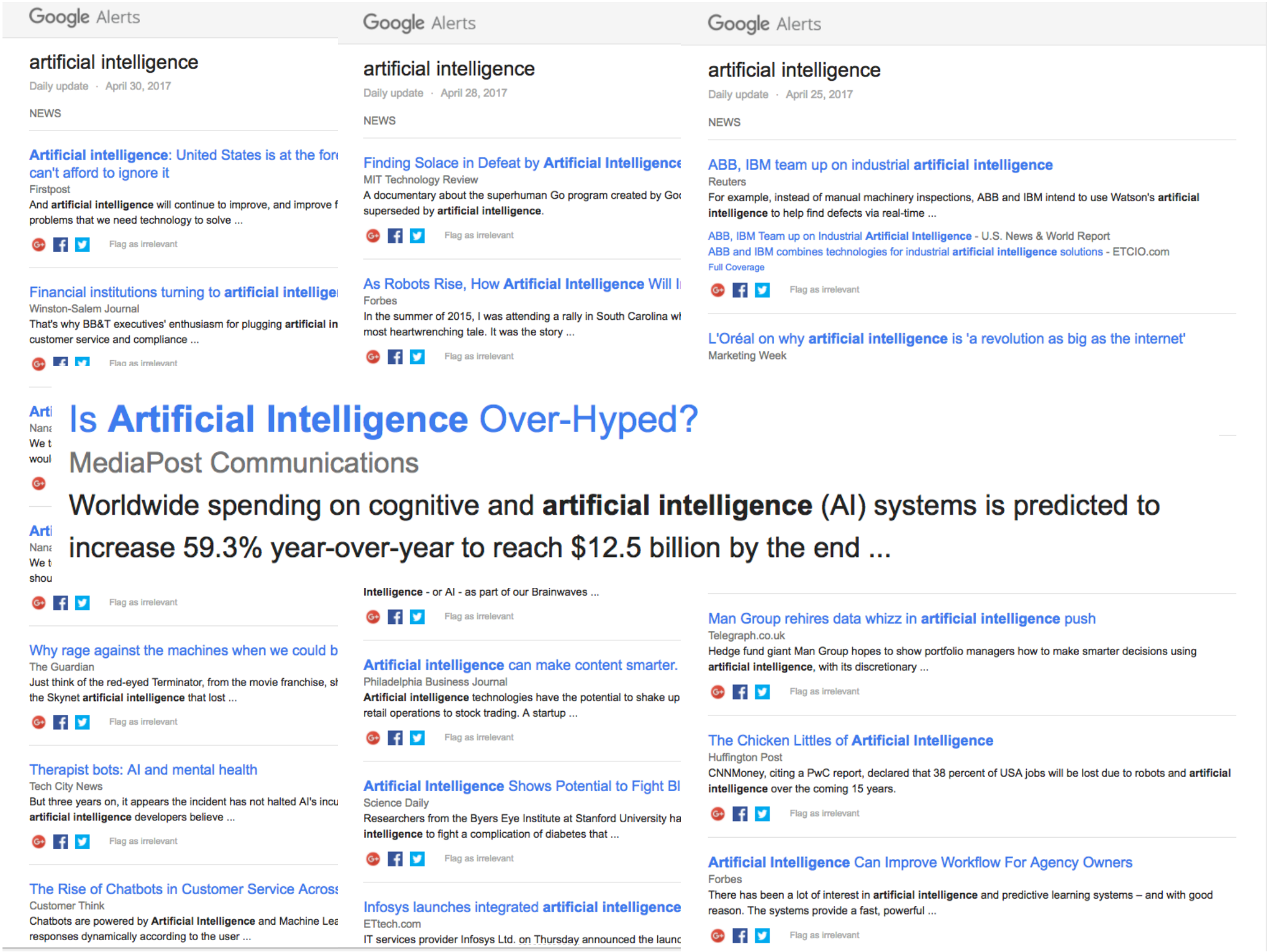
In this thought experiment, a person in the "Chinese room" is passed questions from outside the room, and consults a library of books to formulate an answer







"Physicists and Philosophers united against AI"?



Google Alerts

artificial intelligence

Daily update · April 30, 2017

NEWS

Artificial intelligence: United States is at the for
can't afford to ignore it
Firstpost
And **artificial intelligence** will continue to improve, and improve f
problems that we need technology to solve ...

Flag as irrelevant

Financial institutions turning to **artificial intellige**
Winston-Salem Journal
That's why BB&T executives' enthusiasm for plugging **artificial in**
customer service and compliance ...

Flag as irrelevant

Arti
Nana
We t
woul

Flag as irrelevant

Arti
Nana
We t
shou

Flag as irrelevant

Why rage against the machines when we could b
The Guardian
Just think of the red-eyed Terminator, from the movie franchise, sh
the Skynet **artificial intelligence** that lost ...

Flag as irrelevant

Therapist bots: AI and mental health
Tech City News
But three years on, it appears the incident has not halted AI's incu
artificial intelligence developers believe ...

Flag as irrelevant

The Rise of Chatbots in Customer Service Across
Customer Think
Chatbots are powered by **Artificial Intelligence** and Machine Lea
responses dynamically according to the user ...

Google Alerts

artificial intelligence

Daily update · April 28, 2017

NEWS

Finding Solace in Defeat by **Artificial Intelligence**
MIT Technology Review
A documentary about the superhuman Go program created by Goo
superseded by **artificial intelligence**.

Flag as irrelevant

As Robots Rise, How **Artificial Intelligence** Will I
Forbes
In the summer of 2015, I was attending a rally in South Carolina wh
most heartwrenching tale. It was the story ...

Flag as irrelevant

Is **Artificial Intelligence** Over-Hyped?
MediaPost Communications
Worldwide spending on cognitive and **artificial intelligence** (AI) systems is predicted to increase 59.3% year-over-year to reach $12.5 billion by the end ...

Intelligence - or AI - as part of our Brainwaves ...

Flag as irrelevant

Artificial intelligence can make content smarter.
Philadelphia Business Journal
Artificial intelligence technologies have the potential to shake up
retail operations to stock trading. A startup ...

Flag as irrelevant

Artificial Intelligence Shows Potential to Fight Bl
Science Daily
Researchers from the Byers Eye Institute at Stanford University ha
intelligence to fight a complication of diabetes that ...

Flag as irrelevant

Infosys launches integrated **artificial intelligence**
ETtech.com
IT services provider Infosys Ltd. on Thursday announced the launc

Google Alerts

artificial intelligence

Daily update · April 25, 2017

NEWS

ABB, IBM team up on industrial **artificial intelligence**
Reuters
For example, instead of manual machinery inspections, ABB and IBM intend to use Watson's **artificial intelligence** to help find defects via real-time ...

ABB, IBM Team up on Industrial **Artificial Intelligence** - U.S. News & World Report
ABB and IBM combines technologies for industrial **artificial intelligence** solutions - ETCIO.com
Full Coverage

Flag as irrelevant

L'Oréal on why **artificial intelligence** is 'a revolution as big as the internet'
Marketing Week

Man Group rehires data whizz in **artificial intelligence** push
Telegraph.co.uk
Hedge fund giant Man Group hopes to show portfolio managers how to make smarter decisions using **artificial intelligence**, with its discretionary ...

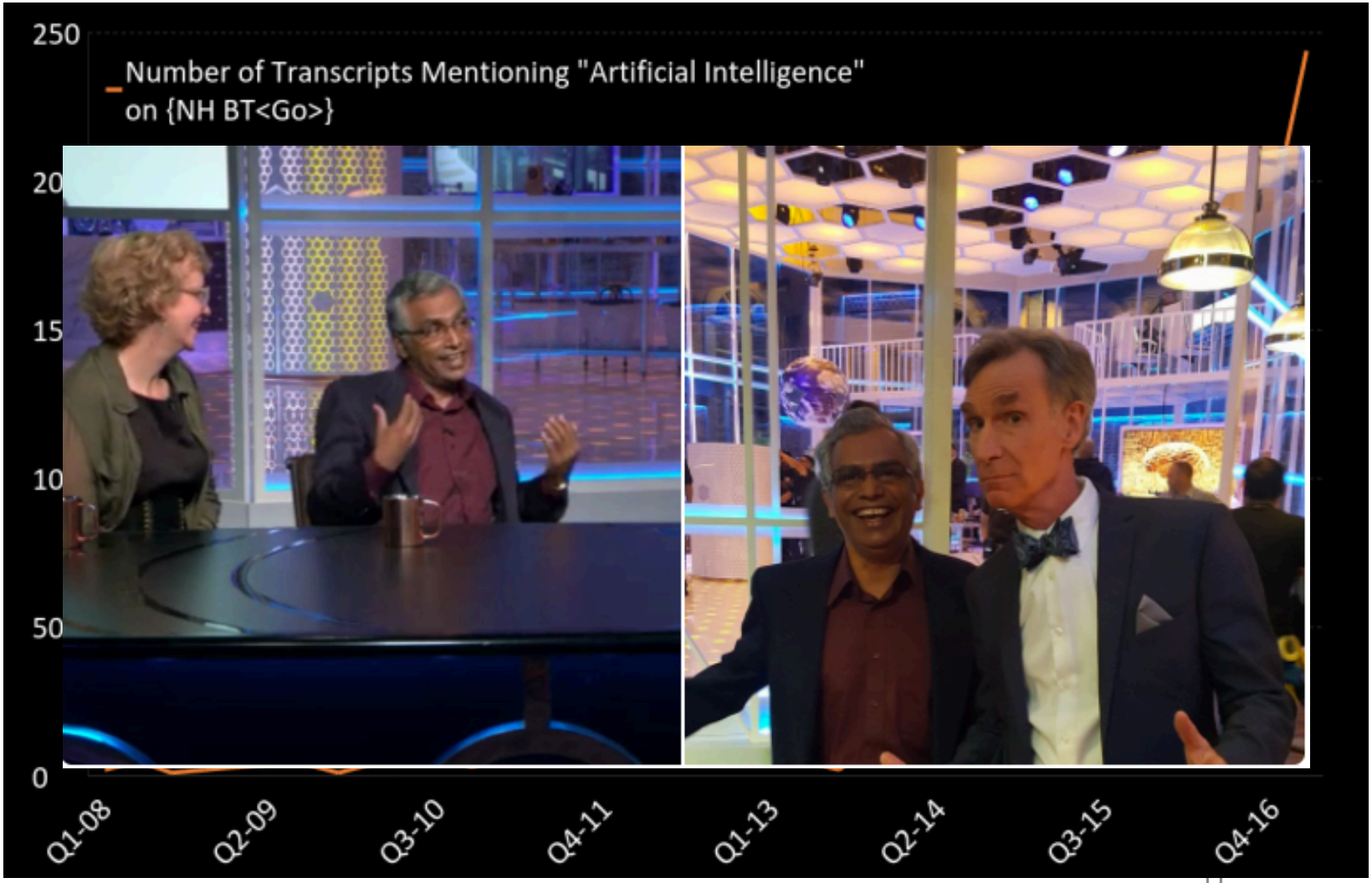
Flag as irrelevant

The Chicken Littles of **Artificial Intelligence**
Huffington Post
CNNMoney, citing a PwC report, declared that 38 percent of USA jobs will be lost due to robots and **artificial intelligence** over the coming 15 years.

Flag as irrelevant

Artificial Intelligence Can Improve Workflow For Agency Owners
Forbes
There has been a lot of interest in **artificial intelligence** and predictive learning systems – and with good reason. The systems provide a fast, powerful ...

Flag as irrelevant


250
Number of Transcripts Mentioning "Artificial Intelligence"
on {NH BT<Go>}
20
15
10
50
0
Q1-08
Q2-09
Q3-10
Q4-11
Q1-13
Q2-14
Q3-15
Q4-16

Bloomberg

The Many Intelligences..

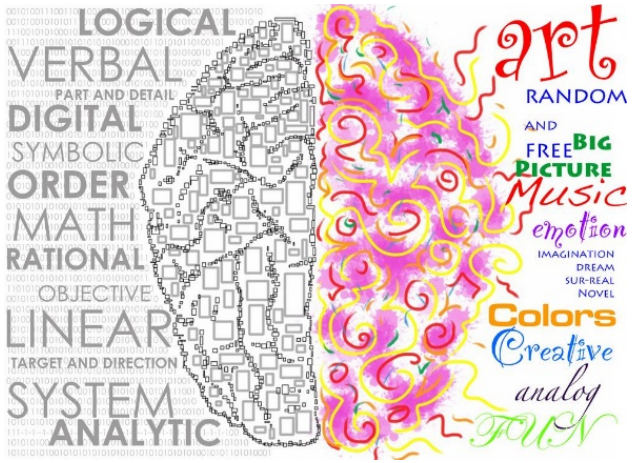
Humans

- Perceptual & Manipulation intelligence that seem to come naturally to us
  - Form the basis for the Captchas..
    - But rarely form the basis for our own judgements about each other's intelligence
- Emotional Intelligence
- Social Intelligence
- Cognitive/reasoning tasks
  - That seem to be what we get tested in in SAT etc.

AI's progress towards intelligence

- 80's --- Expert systems
  - Rule-based systems for many businesses
- 90's -- Reasoning systems
  - Dethroned Kasparov
- 00's: Perceptual tasks
  - Speech recognition common place!
  - Image recognition has improved significantly
- Current: Connecting reasoning and perception

The Many Intelligences..

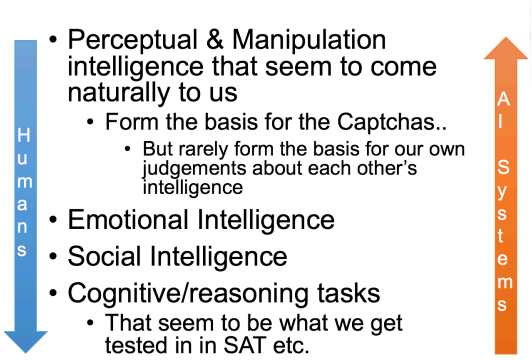


Humans

- Perceptual & Manipulation intelligence that seem to come naturally to us
  - Form the basis for the Captchas..
    - But rarely form the basis for our own judgements about each other's intelligence
- Emotional Intelligence
- Social Intelligence
- Cognitive/reasoning tasks
  - That seem to be what we get tested in in SAT etc.

AI Systems

Explains a lot!



Why did AI develop this "reverse" way?

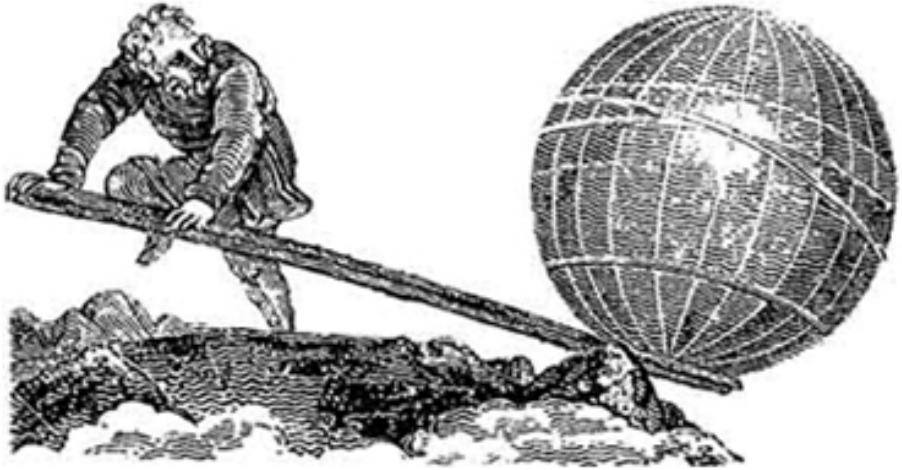
- It is easier to program computers on aspects of intelligence for which we have conscious theories!
  - Ergo the progress in reasoning/cognitive intelligence
- We are not particularly conscious of perceptual (and manipulative) intelligence
  - We had to depend on making machines *learn* the way we had to..
    - **Learn** from data/demonstrations…

Why did AI catch public imagination now?

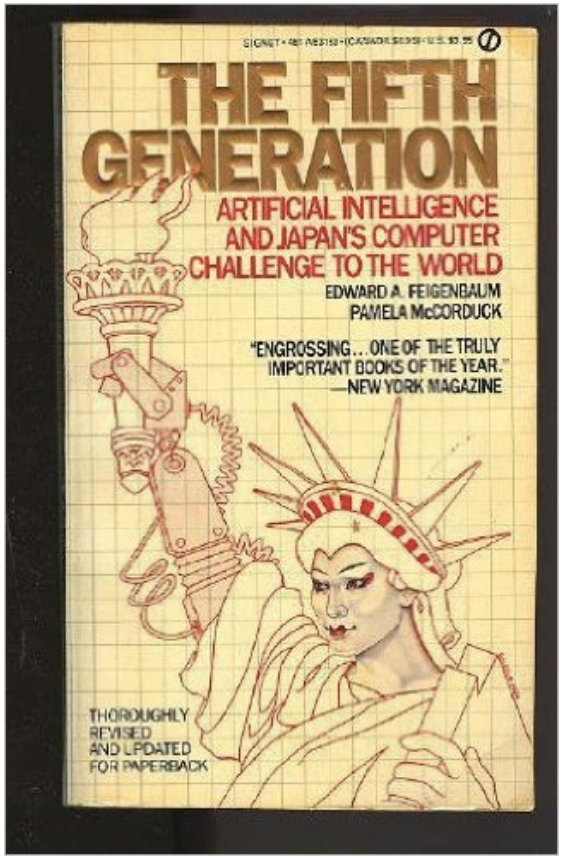
- Early AI was a blind and deaf Socrates
- Perceptual abilities allowed AI to come to all of us
  - On our cell phones; Alexas; Teslas,
- …and now, people suddenly see AI everywhere
  - .. Which also leads to many misperceptions in the public

Are we done?

Irrational Exuberance



If you give me a lever, and a place to stand, I can move the world



Give me a big enough GPU,
large enough data set, and deep enough Network, I will create you super intelligence..

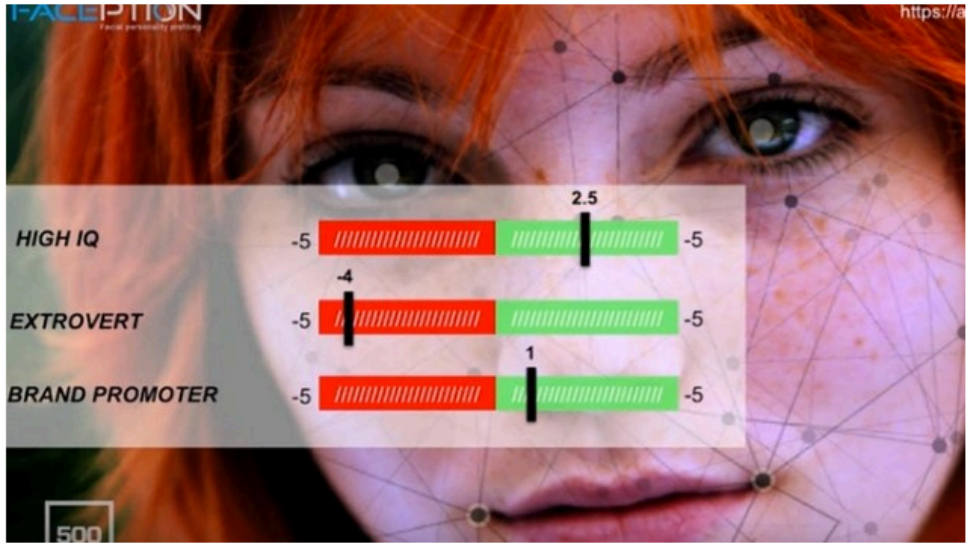
Faception 'can match an individual with various personality traits and types with a high level of accuracy'

New Israeli facial imaging claims to identify terrorists and pedophiles

Tel Aviv start-up Faception says its face 'classifiers' can spot criminals and even great poker players in a split second, but the experts are not convinced

BY SUE SURKES | May 24, 2016, 10:52 pm | 19

Share 2.7K Tweet G+1 5 Email Print Share



An image taken from a May 2016 presentation by Faception co-founder Shai Gilboa (screen capture: YouTube)

NEWSROOM

Email the Newsroom
Facebook
Twitter

A Tel-Aviv based start-up company says it has developed a program to identify personality types such as terrorists, pedophiles, white collar offenders and even great poker players from facial analysis that takes just a fraction of a second.

Get the Start-Up Israel's daily newsletter and never miss our top stories FREE SIGN UP!

Classifiers	CNN	SVM	KNN	LR
AUC	0.9540	0.9303	0.8838	0.8666

Table 1. The AUC results for the four tested face classifiers on criminality.

(b) Three samples in non-criminal ID photo set S_n

Figure 1. Sample ID photos in our data set.

Troubling Study Says Artificial Intelligence Can Predict Who Will Be ...
https://theintercept.com/.../troubling-study-says-artificial-intelligence-can-predict-who...
Nov 18, 2016 - ... Can **Predict** Who Will Be **Criminals Based on Facial Features** ... In a paper titled "Automated Inference on **Criminality using** Face Images," two ...

A New Program Judges If You're a Criminal From Your Facial Features ...
https://motherboard.vice.com/.../new-program-decides-criminality-from-facial-feature...
Nov 18, 2016 - In their paper 'Automated Inference on Criminality **using Face** ... "Also, we find some discriminating structural **features** for **predicting criminality**, ...

Follow

This paper is the exact reason why we need to think about ethics in AI. arxiv.org/abs/1611.04135

2:57 PM - 17 Nov 2016

746 683

Reilly Top 10 | Predicting Criminality
reillytop10.com/2016/12/14/predicting-criminality/
Dec 14, 2016 - Can a computer **predict** who will commit a **crime**? ... disprove the idea that there could be a link between **facial features** and **criminality**. ... Automated inference on **criminality using** face images (Original research article, 2016)

THEMATIC PILLARS

1. Safety-critical AI

Advances in AI have the potential to improve outcomes, enhance quality, and reduce costs in such safety-critical areas as healthcare and transportation. Effective and careful applications of pattern recognition, automated decision making, and robotic systems show promise for enhancing the quality of life and preventing thousands of needless deaths.

However, where AI tools are used to

2. Fair, Transparent, and Accountable AI

AI has the potential to provide societal value by recognizing patterns and drawing inferences from large amounts of data. Data can be harnessed to develop useful diagnostic systems and recommendation engines, and to support people in making breakthroughs in such areas as biomedicine, public health, safety, criminal justice, education, and sustainability.

3.Collaborations between people and AI systems

A promising area of AI is the design of systems that augment the perception, cognition, and problem-solving abilities of people.

Examples include the use of AI technologies to help physicians make more timely and accurate diagnoses and assistance provided to drivers of cars to help them to avoid dangerous situations and crashes.

4. AI, labor and the economy

AI advances will undoubtedly have multiple influences on the distribution of jobs and nature of work. While advances promise to inject great value into the economy, they can also be the source of disruptions as new kinds of work are created and other types of work become less needed due to automation.

Discussions are rising on the best approaches to minimizing potential disruptions, making sure that the fruits of AI advances are widely shared, and competition and innovation is encouraged and not stifled. We seek to study and understand best paths forward, and play a role in this discussion.

5. Social and societal influences of AI

AI advances will touch people and society in numerous ways, including potential influences on privacy, democracy, criminal justice, and human rights. For example, while technologies that personalize information and that support people with recommendations can provide people with valuable assistance, they could also inadvertently or deliberately manipulate and influence opinions.

We seek to promote thoughtful collaboration and open dialogue about the potential subtle and salient influences of AI on people and society.

6. AI for social good

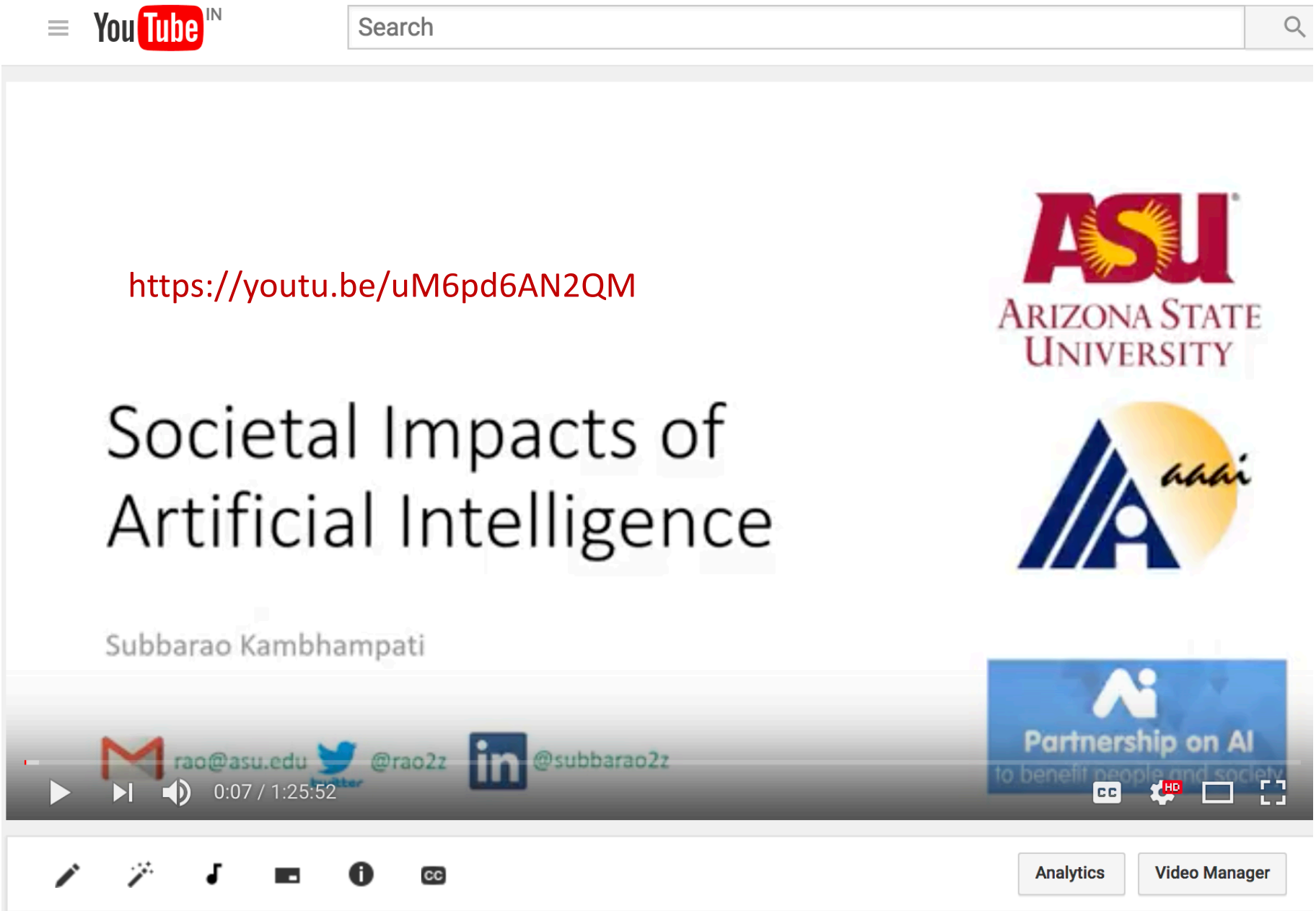
AI offers great potential for promoting the public good, for example in the realms of education, housing, public health, and sustainability. We see great value in collaborating with public and private organizations, including academia, scientific societies, NGOs, social entrepreneurs, and interested private citizens to promote discussions and catalyze efforts to address society's most pressing challenges.

Some of these projects may address deep societal challenges and will be moonshots - ambitious big bets that could have far-reaching impacts. Others may be creative ideas that could quickly produce positive results by harnessing AI advances.

7. Special initiatives

Beyond the specified thematic pillars, we also seek to convene and support projects that resonate with the tenets of our organization. We are particularly interested in supporting people and organizations that can benefit from the Partnership's diverse range of Partners.

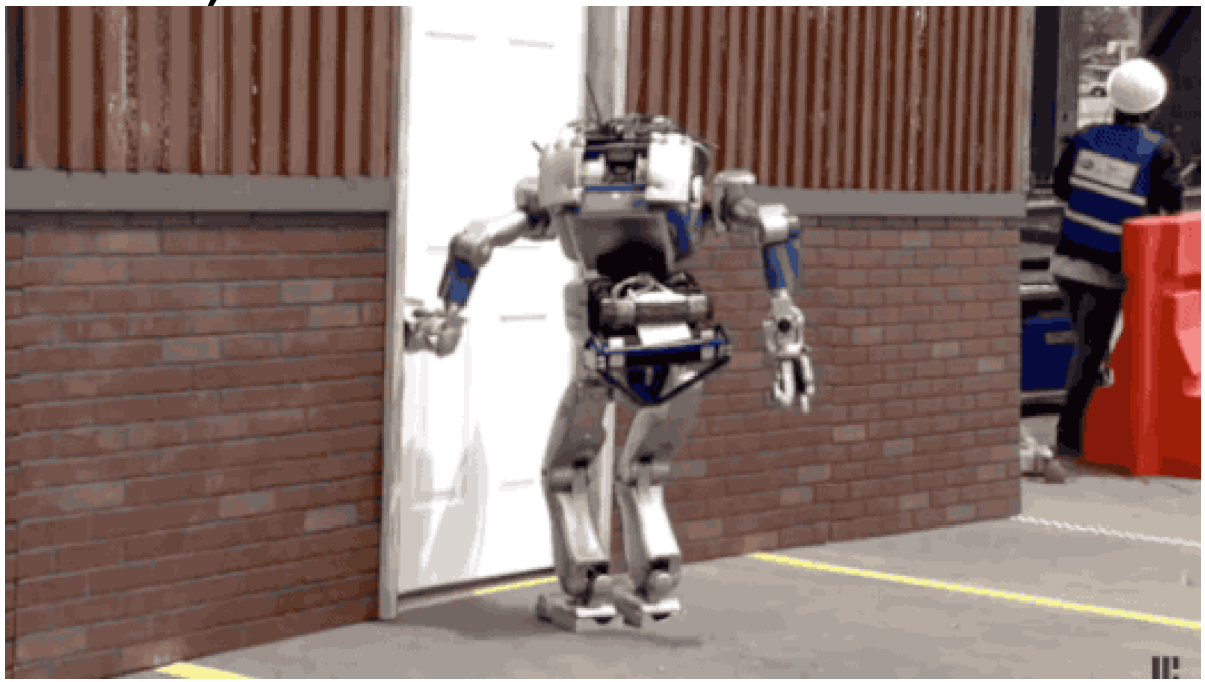
We are open-minded about the forms that these efforts will take.


YouTube IN
Search
https://youtu.be/uM6pd6AN2QM
ASU
ARIZONA STATE UNIVERSITY
Societal Impacts of Artificial Intelligence
aaai
Subbarao Kambhampati
rao@asu.edu
@rao2z
@subbarao2z
Partnership on AI
to benefit people and society
0:07 / 1:25:52
Analytics
Video Manager
Societal Impacts of Artificial Intelligence
Subbarao Kambhampati
Channel settings
137 views

Thresholds

("You have come a long way, Robbie!
But boy do you have a long ways still to go…")

- (Knowledge-based) Learning from fewer examples
- Commonsense
- Incompleteness
- Interaction (with humans)



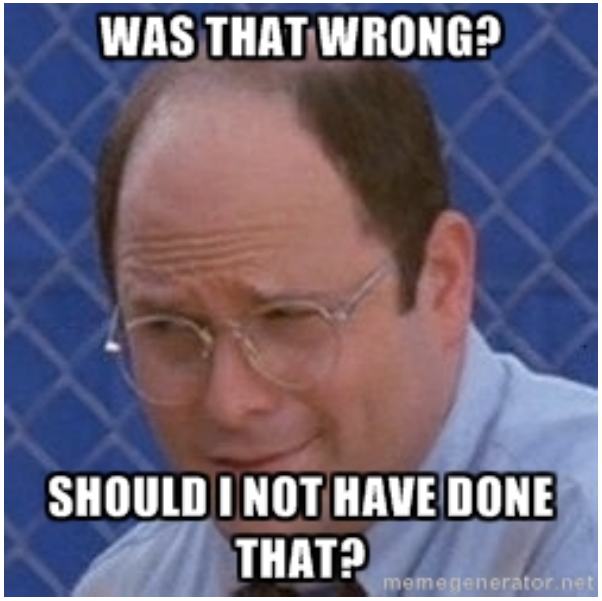
Still Elusive Commonsense

- "Commonsense" elaborates partial specifications of facts, observations, norms, goals….
  - Which trip did Magellan Die?
- Winograd Schema Challenge
  - The women stopped taking pills because they were pregnant
  - The women stopped taking pills because they were carcinogenic

The world is full of obvious things that nobody by any chance ever observes

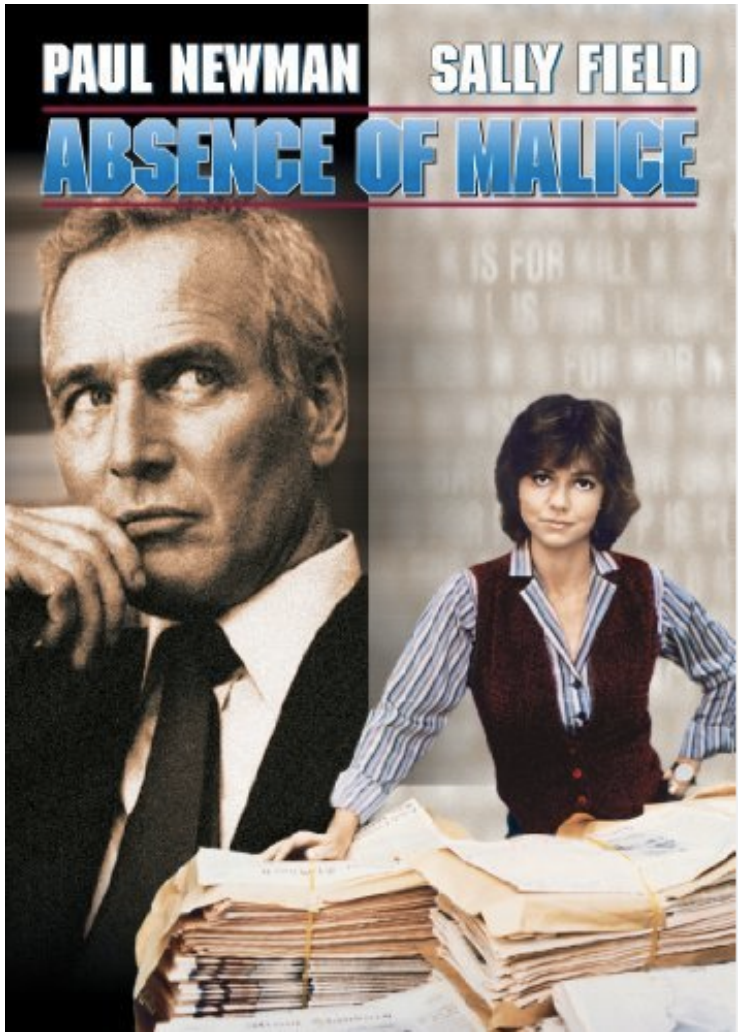
--Christopher in the "Curious incident of the dog in the night time"

(Inadvertently channeling Sherlock Holmes/ Sir Arthur Conon Doyle)



You can cause more destruction with ignorance without any malice..

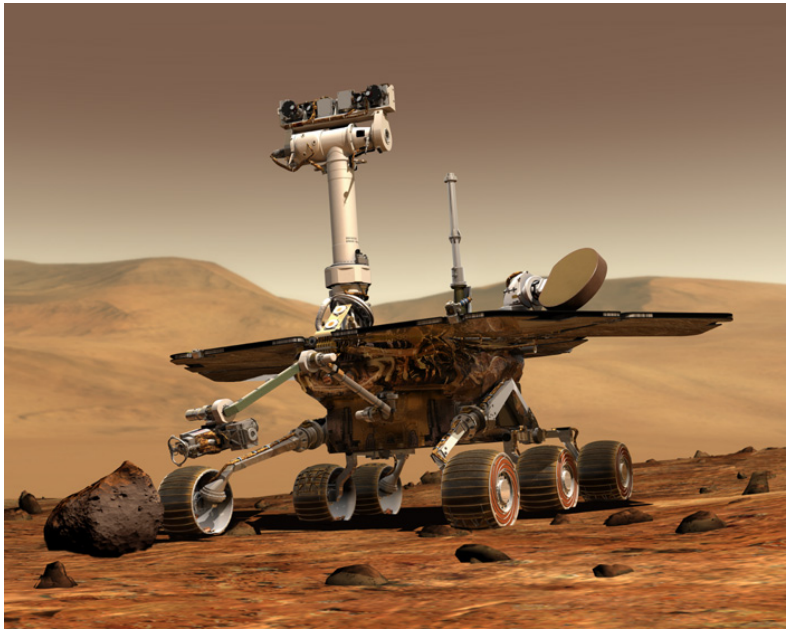
- Much of the knowledge of the agents is going to be incomplete
  - Both the world dynamics and objectives

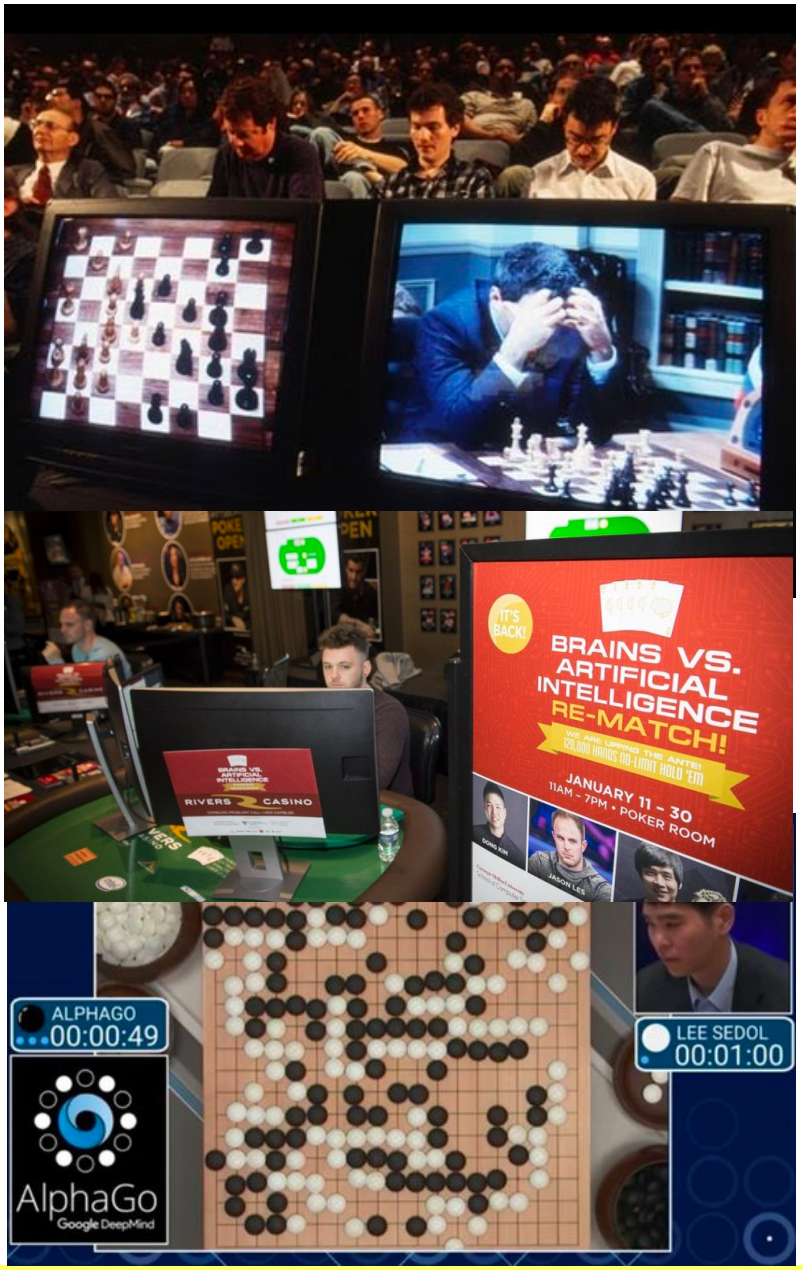


Won’t somebody please think of the Humans?

AI's Curious Ambivalence to humans..

- Our systems seem happiest
  - either far away from humans
  - or in an adversarial stance with humans





You want to help humanity, it is the people that you just can't stand…

IJCAI-16

25th International Joint Conference on Artificial Intelligence

New York City, July 9–15, 2016
www.ijcai-16.org

Special Theme: Human Aware AI

Conference Chair	Program Chair	Local Arrangements Committee Chair	IJCAI Secretary-Treasurer	IJCAI Executive Secretary
Gerhard Brewka	**Subbarao Kambhampati**	**Ernest Davis**	**Bernhard Nebel**	**Vesna Sabljakovic-Fritz**
Leipzig University, Germany	Arizona State University, Tempe	New York University	Albert-Ludwigs-Universität Freiburg	Vienna University of Technology, Austria

Organizing Institutions

IJCAI The International Joint Conferences on Artificial Intelligence

AAAI The Association for the Advancement of Artificial Intelligence

Why intentionally design a dystopian future and spend time being paranoid about it?

PCWorld FROM IDG

NEWS | REVIEWS | HOW-TO | VIDEO | BUSINESS | LAPTOPS | TABLETS | PHONES | HARDWARE | SECURITY | SOFTWARE | GADGETS

Home / Analytics

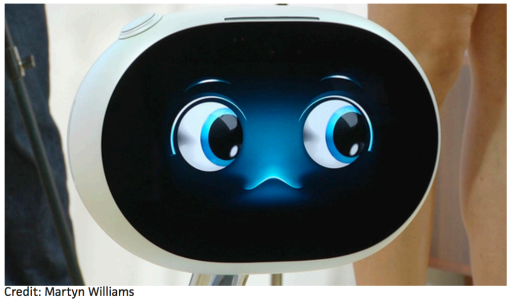
NEWS

How 'human-aware' AI could save us from the robopocalypse

AI should relate to people as an apprentice, not a tool, one researcher says

By Katherine Noyes
Senior U.S. Correspondent, IDG News Service | JUL 13, 2016 10:05 AM PT

bold360 Engage Your Customers at the Critical Moments in Their Journey LEARN MORE
Customer Engagement in a Connected World | Modern Consumers Need an Omni-Channel Strategy | Playbook to Activating Customer Engagement | Turn Browsers Into Buyers, Drive Repeat Business



Credit: Martyn Williams

MORE LIKE THIS

AI + humans = kick-ass cybersecurity

The future of artificial intelligence: Computers will take your job

Can robots make art? Yes, but don't ask them to write a poem

VIDEO
AMD Radeon Vega Frontier Edition Hands-on

8.2 Recommendations

JASON offers the following recommendations to DoD senior leadership:

1. DoD should both track (via a knowledgeable cadre) and invest in (via a 6.1 research portfolio) the most dynamic and rapidly advancing areas of AI, including, but by no means limited to DL.

55

JASON Briefing on "The Path to General AI goes through Human-Aware AI"; June 2016

2. DoD should support the development of a discipline of AI engineering, accelerating the progress of the field through Shaw's "craft" and (empirical) "commercial" stages. A particular focus should be advancing the "illities" in support of DoD missions.
3. DoD's portfolio in AGI should be modest and recognize that it is not currently a rapidly advancing area of AI. The field of human augmentation via AI is much more promising and deserves significant DoD support.
4. DoD should support the curation and labeling, for research, of its unique mission-related large data sets. Wherever possible, operational data should be saved for future research use in support of AI for DoD-unique missions.
5. DoD should create and provide centralized resources for its intramural and extramural researchers (MOSIS-like), including labeled data sets and access to large-scale GPU training platforms.
6. DoD should survey the mission space of embedded devices for potential breakthrough applications of AI, and should consider investing in special-purpose accelerators to support AI inference in embedded devices for DoD missions if such applications are identified.

Seeking new algorithms for human-aware AI

Over the years, AI algorithms have become able to solve problems of increasing complexity. However, there is a gap between the capabilities of these algorithms and the usability of these systems by humans. *Human-aware* intelligent systems are needed that can interact intuitively with users and enable seamless machine-human collaborations. Intuitive interactions include shallow interactions, such as when a user discards an option recommended by the system; model-based approaches that take into account the users' past actions; or even deep models of user intent that are based upon accurate human cognitive models. Interruption models must be developed that allow an intelligent system to interrupt the human only when necessary and appropriate. Intelligent systems should also have the ability to augment human cognition, knowing which information to retrieve when the user needs it, even when they have not prompted the system explicitly for that information. Future intelligent systems must be able to account for human social norms and act accordingly. Intelligent systems can more effectively work with humans if they possess some degree of emotional intelligence, so that they can recognize their users' emotions and respond appropriately. An additional research goal is to go beyond interactions of one human and one machine, toward a "systems-of-systems", that is, teams composed of multiple machines interacting with multiple humans.

Human-AI system interactions have a wide range of objectives. AI systems need the ability to represent a multitude of goals, actions that they can take to reach those goals, constraints on those actions, and other factors, as well as easily adapt to modifications in the goals. In addition, humans and AI systems

must share common goals and have a mutual understanding of them and relevant aspects of their current states. Further investigation is needed to generalize these facets of human-AI systems to develop systems that require less human engineering.

Heading toward Artificial Intelligence 2.0

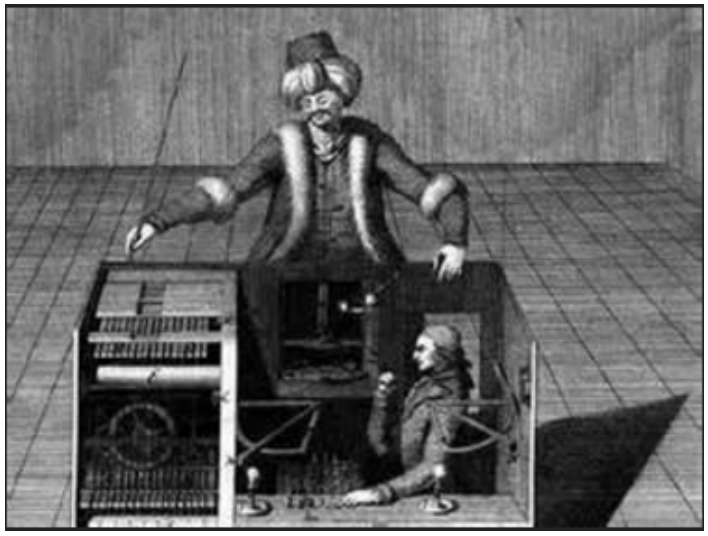
Yunhe Pan

Chinese Academy of Engineering, Beijing 100088, China

AI 2.0 technology will possess distinguishing features, such as the process of combining data-driven and knowledge guidance into autonomous machine learning that is both explainable and more general. In addition, there will be a move away from the processing of categorical data—such as visual, auditory, and written data—and toward cross-media cognition, learning, and inference. Furthermore, there will be a move toward new forms of hybrid-augmented intelligence, from the pursuit of an intelligent machine to high-level human-machine collaboration and fusion. Another area will involve the formation of technologies and platforms to promote crowd-based intelligence built on individual intelligence in order to form a higher level of community intelligence that is based on the Internet. Finally, there will be an extension from research involving robotics to more expansively autonomous-intelligent systems focused on developing intelligent machinery and products.

But isn't this cheating?

- Doesn't putting human in the loop dilute the AI problem?
- Won't it be cheating?
  - Like the original Mechanical Turk…



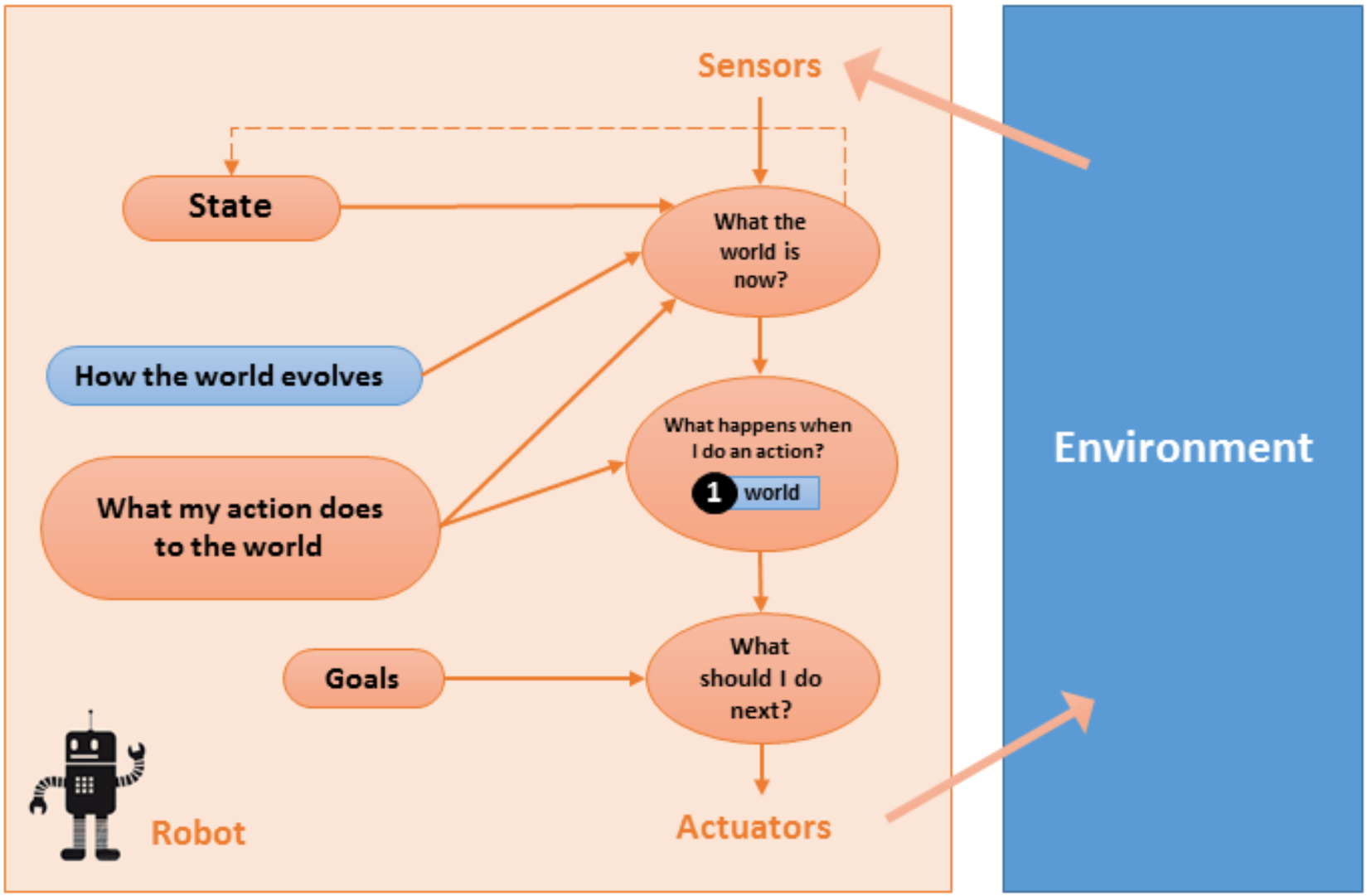
- NO!
  - Expands reach and scope of AI enterprise
  - Reduces some of the off-the-top worries about AI
  - Brings up novel research challenges

Many Intelligences..

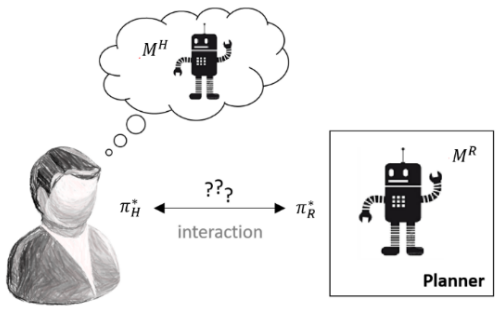
- Perceptual & Manipulation intelligence that seem to come naturally to us
  - Form the basis for the Captchas..
    - But rarely form the basis for our own judgements about each other's intelligence
- Emotional Intelligence
- Social Intelligence

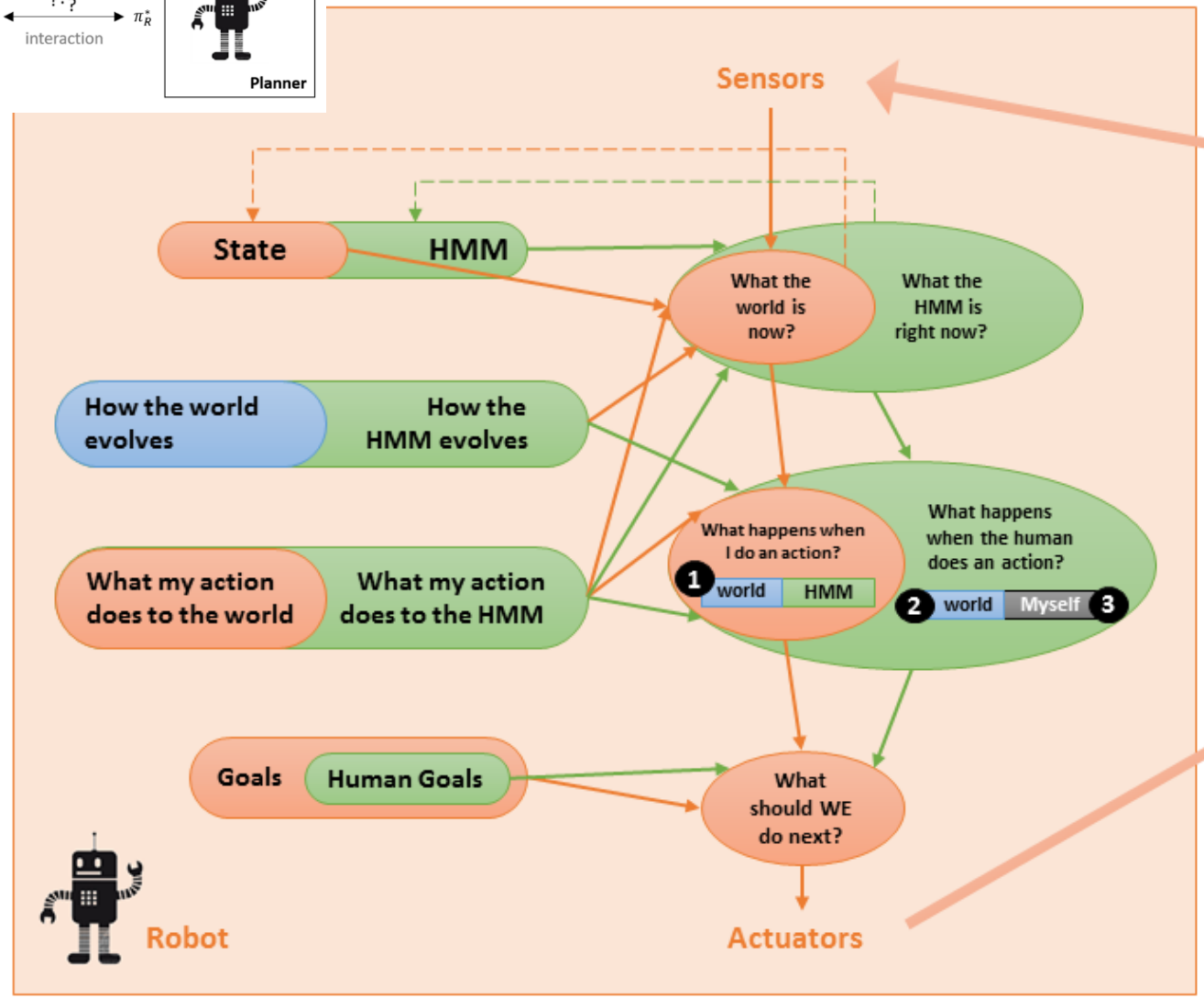
- Cognitive/reasoning tasks
  - That seem to be what we get tested in in SAT etc.

Architecture of an Intelligent Agent



Architecture of an Intelligent Agent teaming with a human

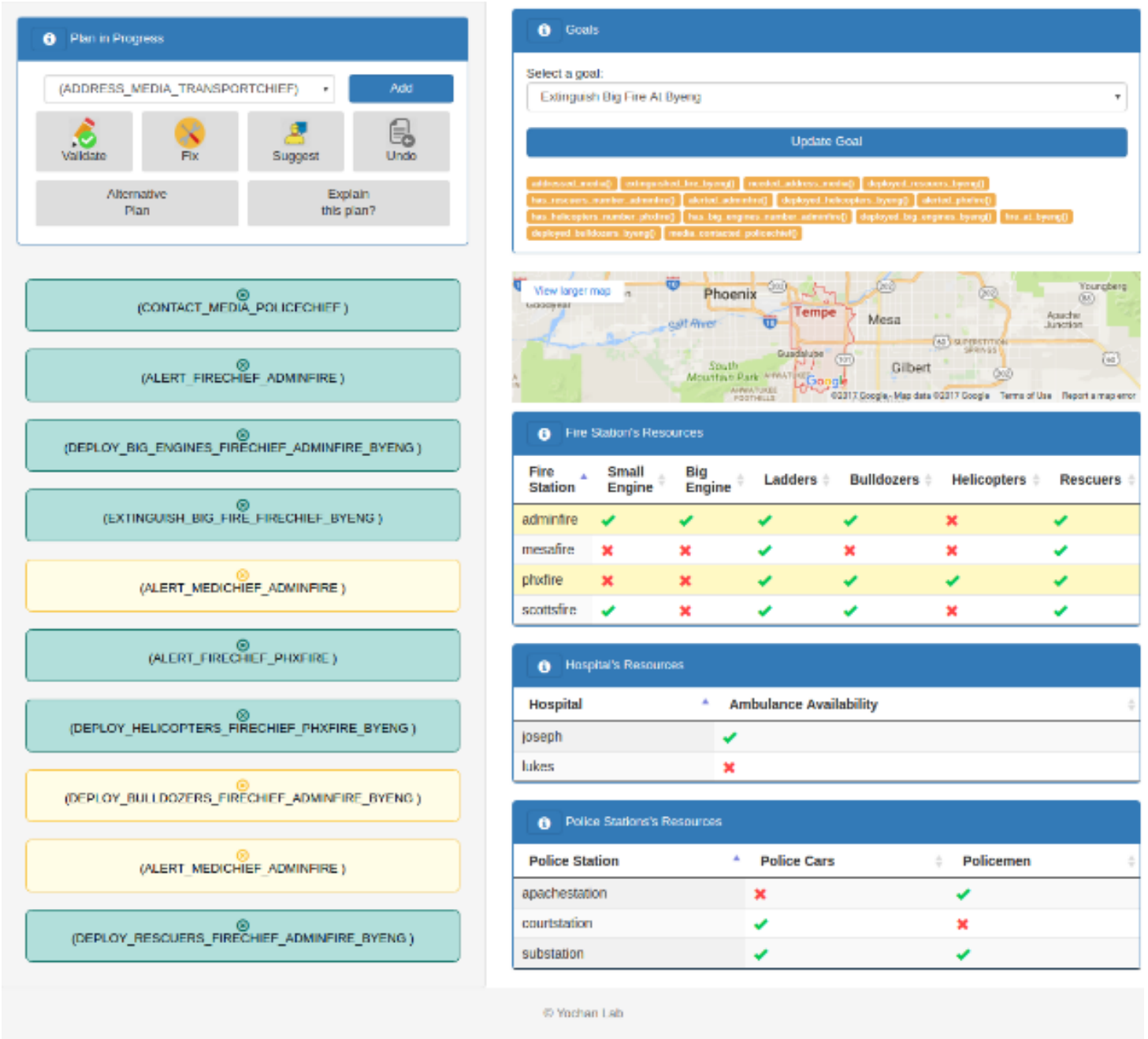




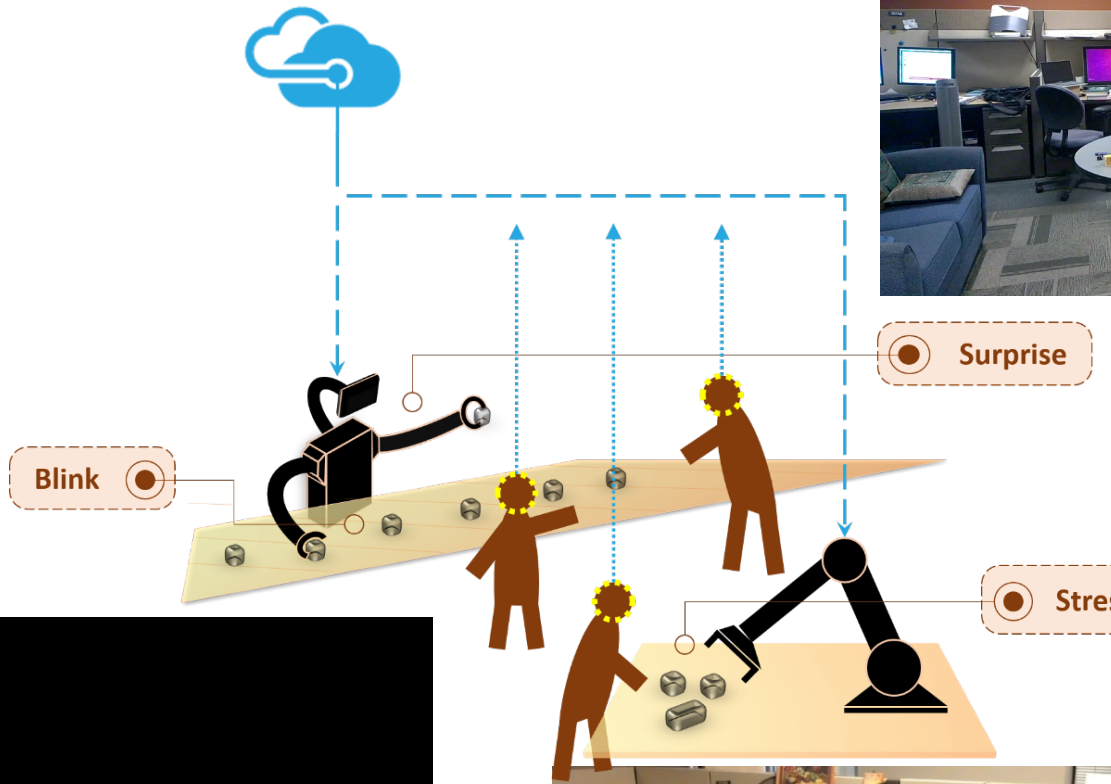
HMM= Human Mental Model

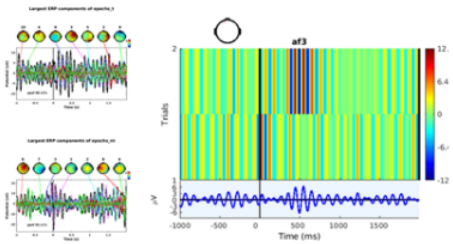
rao@asu.edu @rao2z @subbarao2z

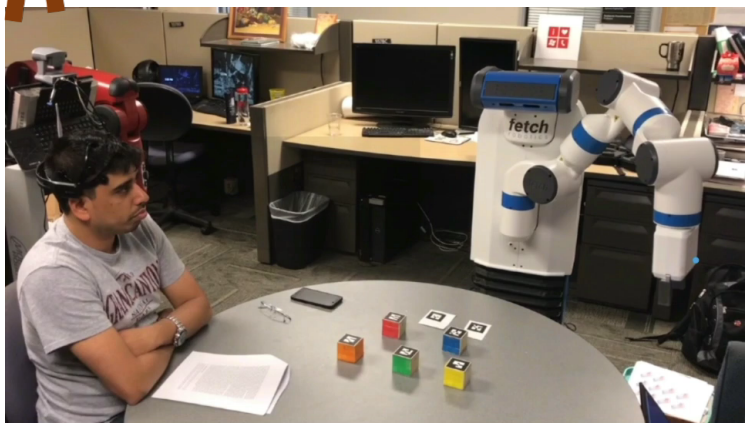
Human-in-the-Loop Planning

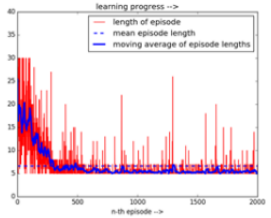


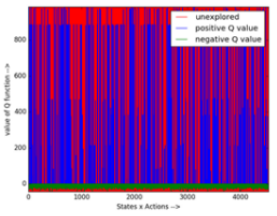
Intention Recognition with Emotive



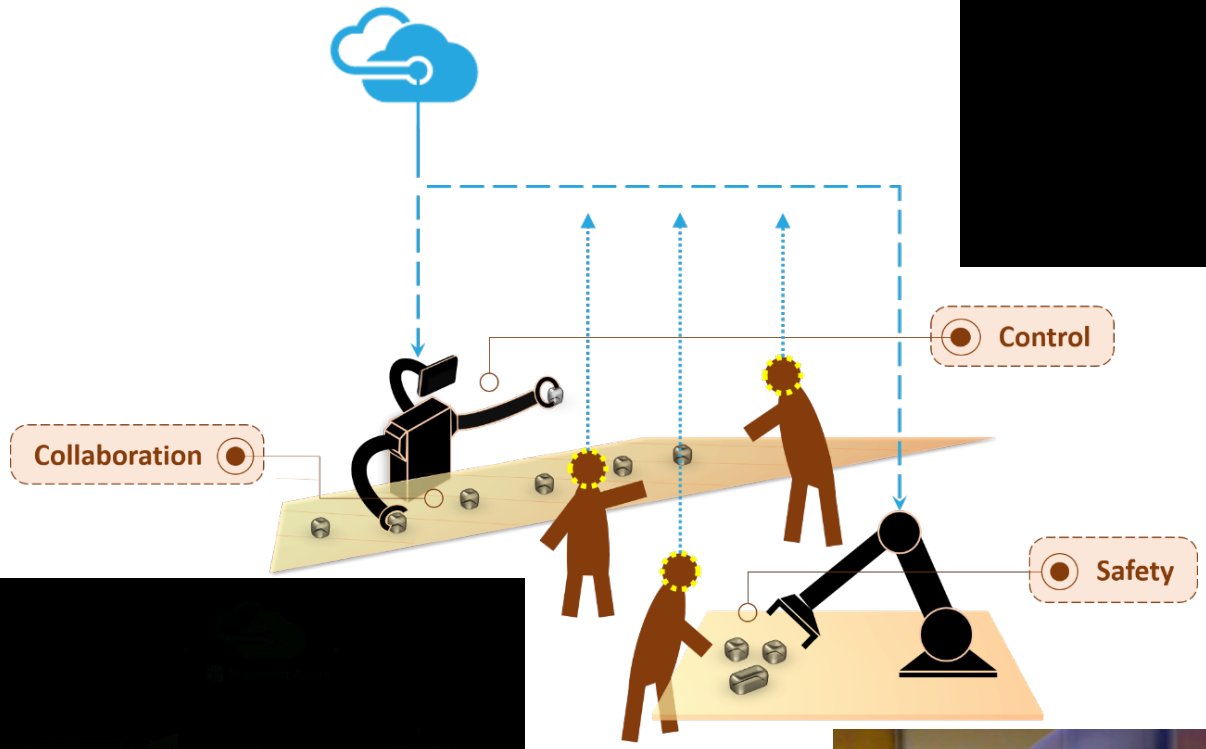


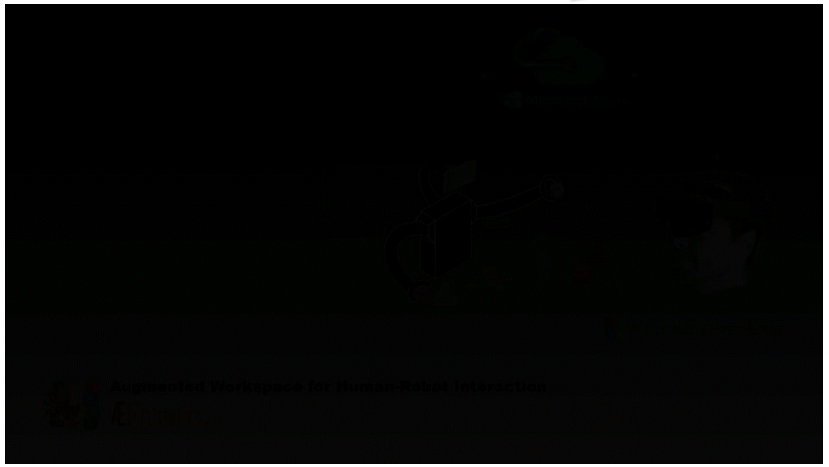




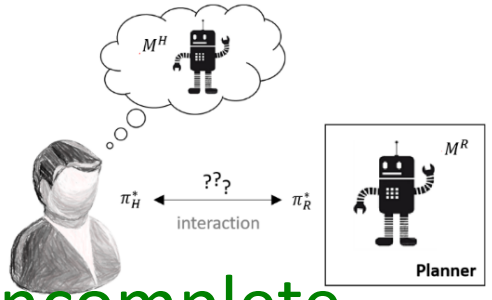


Intention Projection with Hololens





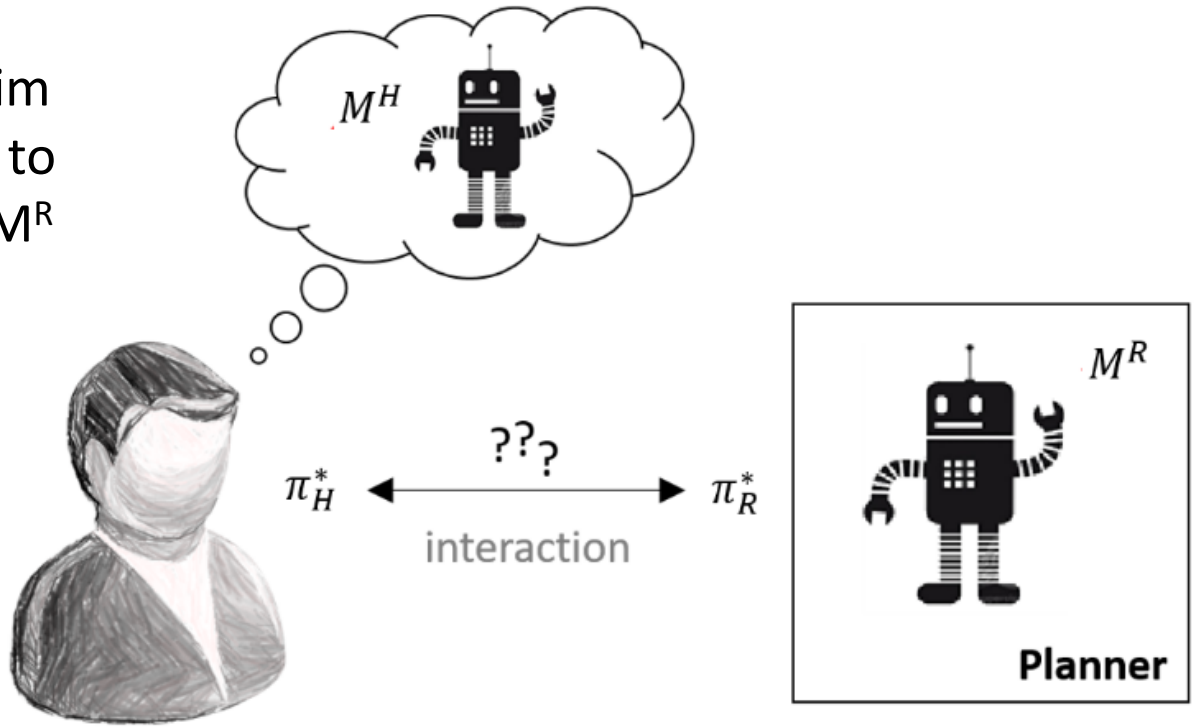
Challenges in Human-Aware Planning



- Interpret what humans are doing based on incomplete human and domain models (Modeling)
  - Plan/goal/intent recognition
- Plan with incomplete domain models (Decision Making)
  - Robust planning/execution support with “lite” models
  - Proactive teaming support
- Explicable Behavior, Explanations/Excuses (Interaction/Communication)
  - How should the human and robot coordinate
- Understand effective interactions between humans and machines (Evaluation)
  - Human factor study

Interaction Requires Modeling the Human

Explicability: Aim to get π^R closer to π^H (by getting M^R closer to $M^{H)}$

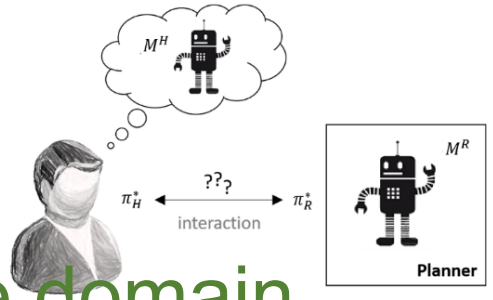


Explanation:

Tell human how to get M^H closer to M^R

--What is the minimum number of changes needed in M^H such that π^R would be optimal plan.

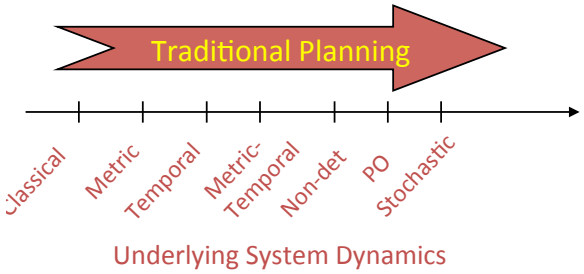
Overview of our work



- How to learn and plan with incomplete domain models
  - Complete--Approximate--Shallow
- How to plan to be useful to the human
  - Avoiding conflicts and offering serendipitous help
- How to make planned behavior explicable or provide explanations to the human in the loop
  - Humans will parse the behavior in terms of their understanding of the Robot's model
- How to recognize and evaluate what are the desiderata for fluent teaming with humans
  - As the "paper clip" assistant shows, we AI'ers are not great at guessing what humans "like" ☹

Overview of our ongoing work

- How to learn and plan with incomplete domain models
  - Complete--Approximate--Shallow
- How to plan to be useful to the human
  - Avoiding conflicts and offering serendipitous help
- How to make planned behavior explicable or provide explanations to the human in the loop
  - Humans will parse the behavior in terms of their understanding of the Robot's model
- How to recognize and evaluate what are the desiderata for fluent teaming with humans
  - As the "paper clip" assistant shows, we AI'ers are not great at guessing what humans "like" ☹

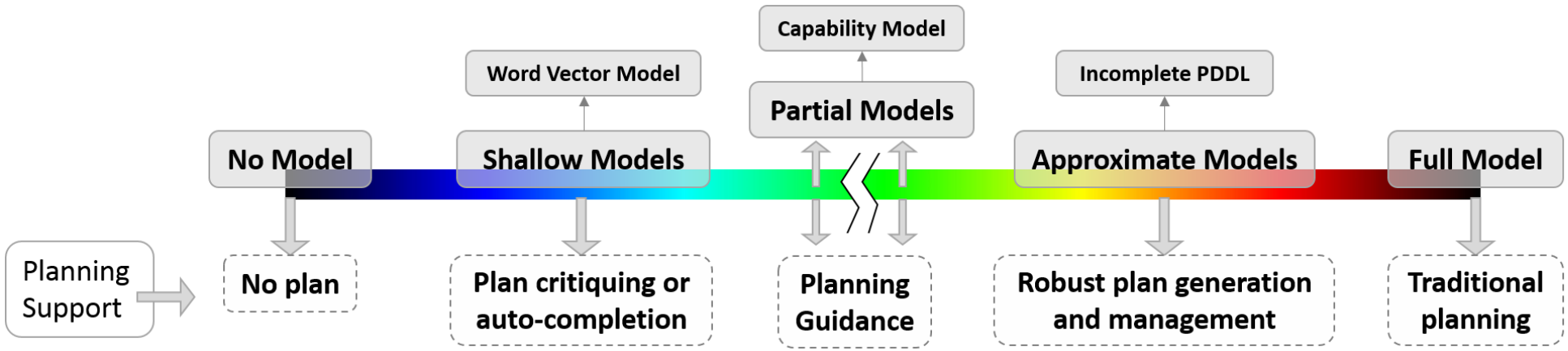


Spectrum of Domain Models

Best Student Paper Nominee

[AAMAS 2016] [AAMAS 2015] [AIJ 2017; ICAPS 2014; IJCAI 2009, 2007]

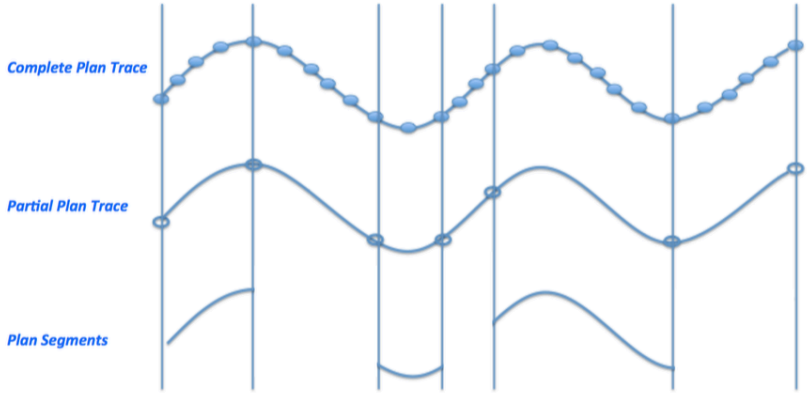
Increasing degree of incompleteness of planning models



← Associative/uninterpretable

Causal/interpretable →

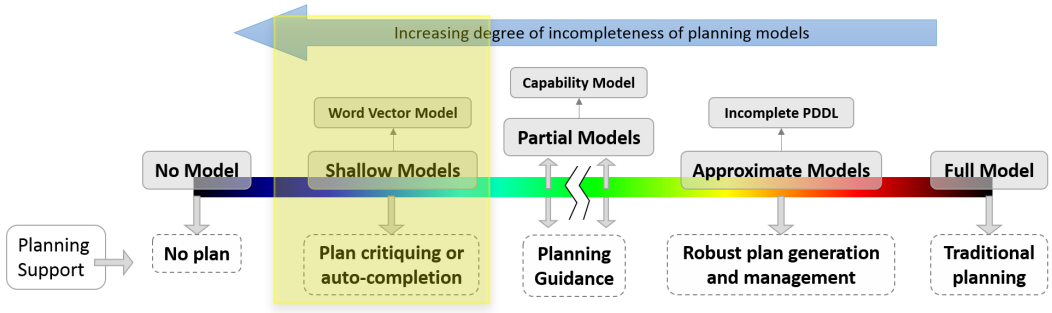
Ease of learning/acquiring the models



Note the contrast to ML research where the progress is going from uninterpretable/non-causal models *towards* interpretable and causal models.
So we might meet in the middle!

Action Vector Models

- View observed action sequences as “sentences” in a language whose “words” are the actions
- Apply skip-gram models to these sequences and embed the action “words” in a higher dimensional space
  - The proximity of the action words in that space is seen as their “affinity”
- Use the action affinities as a way to drive planning and plan recognition



Action Vector Models can be used to Recognize Plans

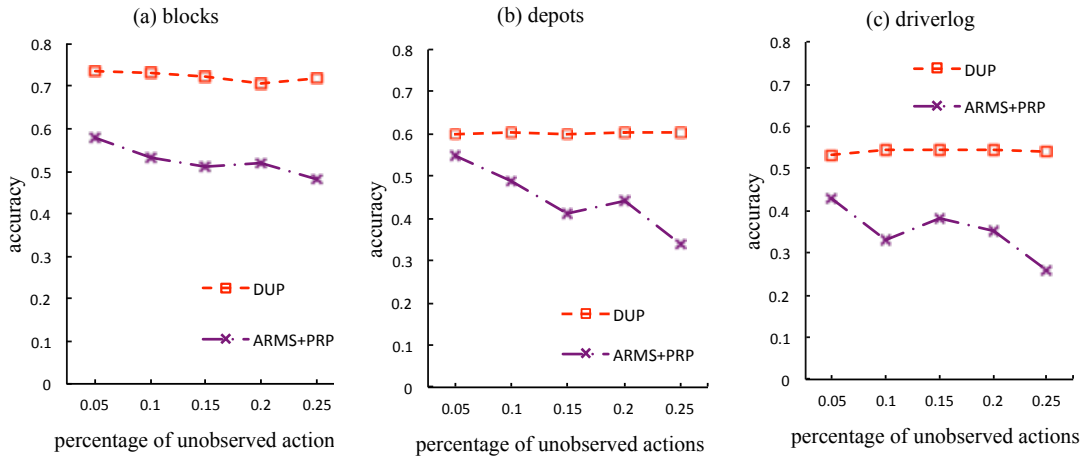
With the learnt vectors w_i, we can predict the target plan (as the most consistent with the affinities). We use an EM procedure to speedup the prediction.

$$\mathcal{F}(\tilde{p}) = \sum_{k=1}^{M} \sum_{-c \le j \le c, j \neq 0} \log p(w_{k+j} | w_k)$$

- M = |the target plan|

The target plan to be recognized

Learning shallow models can avoid overfitting!!



Algorithm 1 Framework of our DUP algorithm

Input: plan library $\mathcal{L}$, observed actions $\mathcal{O}$
Output: plan $\tilde{p}$

1: learn vector representation of actions
2: initialize $\Gamma_{o,k}$ with $1/M$ for all $o \in \bar{\mathcal{A}}$, when k is an unobserved action index
3: **while** the maximal number of repetitions is not reached **do**
4: sample unobserved actions in $\mathcal{O}$ based on Γ
5: update Γ based on Equation (6)
6: project Γ to [0,1]
7: **end while**
8: select actions for unobserved actions with the largest weights in Γ
9: **return** $\tilde{p}$

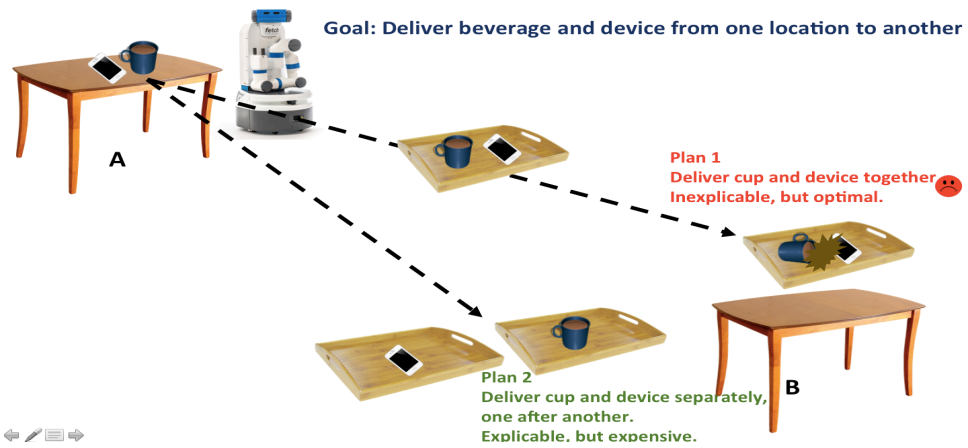
Nominated for Best Student Paper Award at [AAMAS16]

Overview of our ongoing work

- How to learn and plan with incomplete domain models
  - Complete--Approximate--Shallow
- How to plan to be useful to the human
  - Avoiding conflicts and offering serendipitous help
- How to make planned behavior explicable or provide explanations to the human in the loop
  - Humans will parse the behavior in terms of their understanding of the Robot's model
- How to recognize and evaluate what are the desiderata for fluent teaming with humans
  - As the "paper clip" assistant shows, we AI'ers are not great at guessing what humans "like" ☹

When is a plan "Explicable" to the human in the loop?

- The robot generates its plan of action using its model M_R
- The human "interprets" this plan in light of her understanding of the Robot's model M^*_R
- M_R and M^*_R can be quite different..
- Differences can be a result of:
  - Different capabilities (e.g., possible actions)
  - Different knowledge (e.g., level of modeling)
  - Different interpretation of behaviors (e.g., plans) interacting with the world -- more than just trajectory planning!



$$\underset{\pi_{M_R}}{\operatorname{argmin}}\, cost(\pi_{M_R}) + \alpha \cdot dist(\pi_{M_R}, \pi_{\mathcal{M}^*_R})$$

But, alas, M^*_R is not known!

Learning Human Expectation via Explicability Labeling

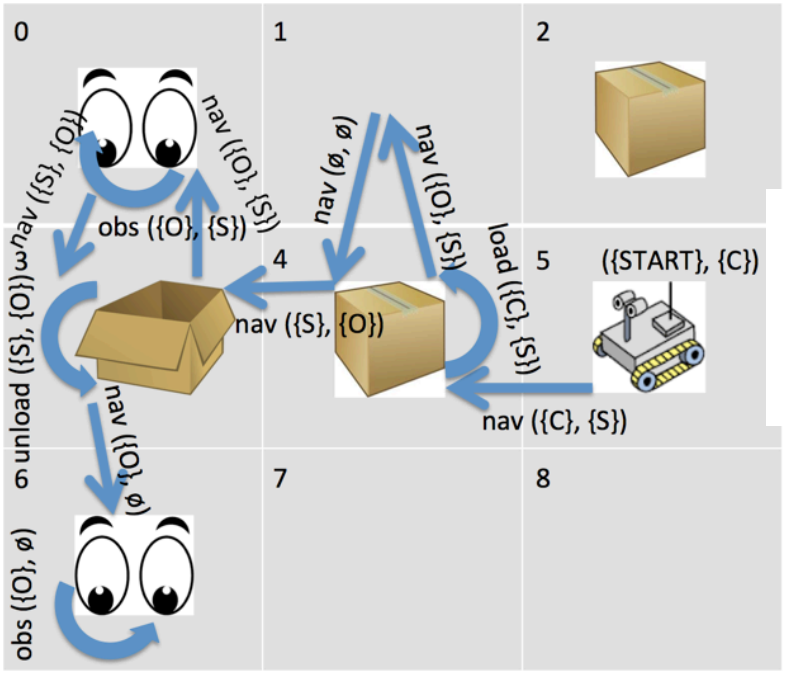
Problem: M^*_R is not known

Solution: Learn it, but indirectly as a labeling scheme..

$$\underset{\pi_{M_R}}{\operatorname{argmin}}\, cost(\pi_{M_R}) + \alpha \cdot dist(\pi_{M_R}, \pi_{\mathcal{M}^*_R})$$

$$dist(\pi_{M_R}, \pi_{\mathcal{M}^*_R}) = F \circ \mathcal{L}^*(\pi_{M_R})$$

$$\underset{\pi_{M_R}}{\operatorname{argmin}}\, cost(\pi_{M_R}) + \alpha \cdot F \circ \mathcal{L}^*_{CRF}(\pi_{M_R} | \{S_i | S_i = \mathcal{L}^*(\pi^i_{M_R})\})$$



Analogy: Think of learning how to write address labels so the postal carrier can understand..

- Task labels (to associate with actions). For example:
  - Collect
  - Store
  - Observe

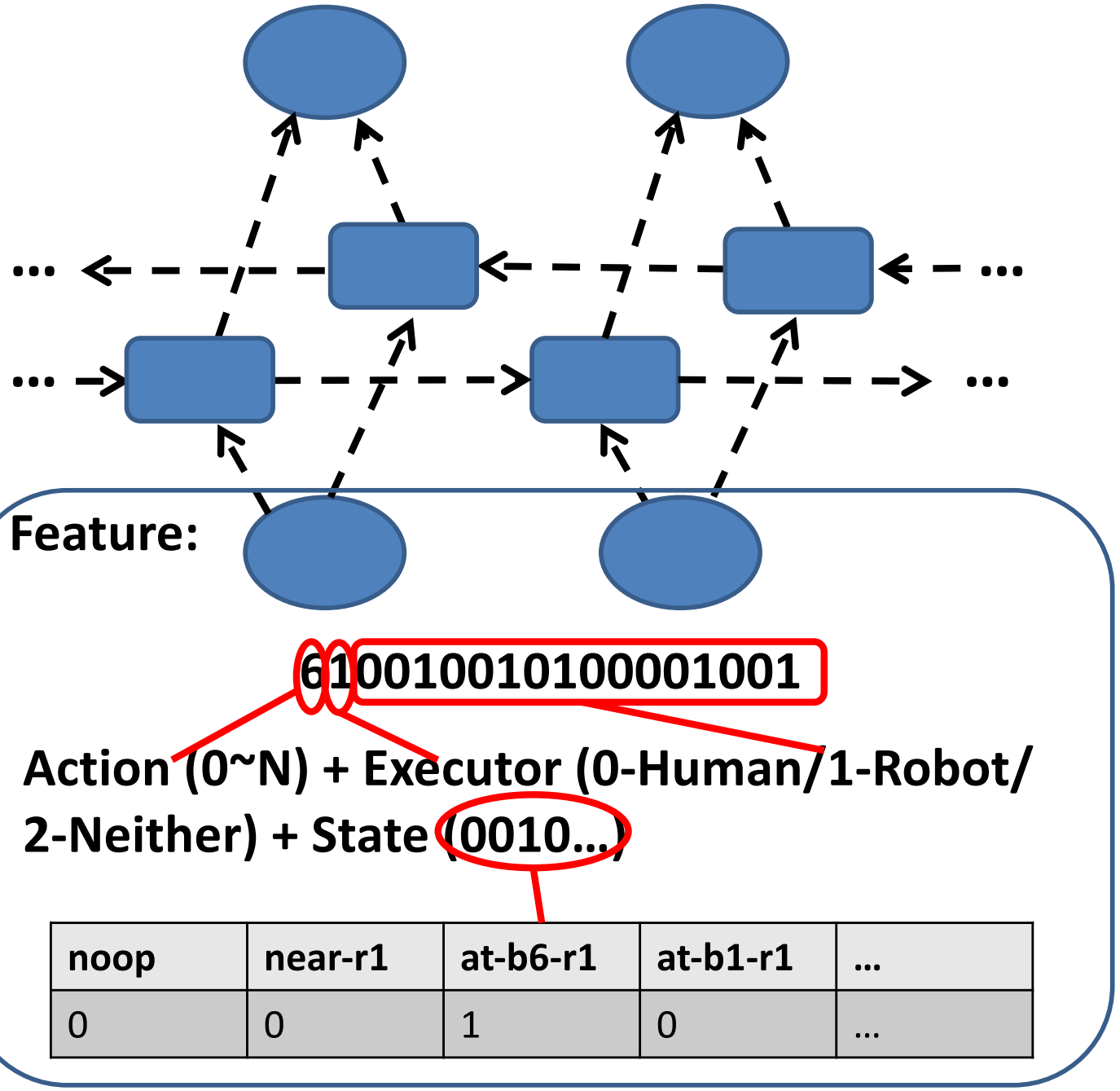
More than one label is allowed for actions

$$\underset{\pi_{M_R}}{\operatorname{argmin}}\, cost(\pi_{M_R}) + \alpha \cdot F \circ \mathcal{L}^*_{CRF}(\pi_{M_R} | \{S_i | S_i = \mathcal{L}^*(\pi^i_{M_R})\})$$

Bi-LSTM as Task Predictor for Plan Explicability

Motivation:

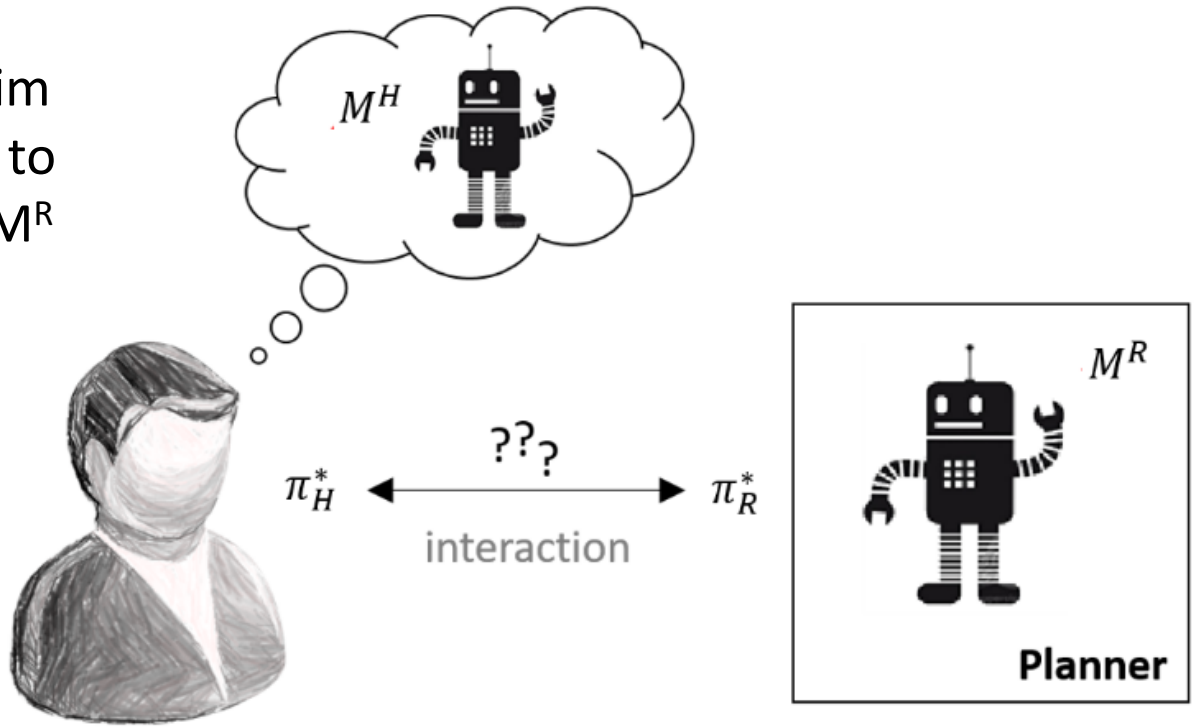
1. **Consider future inputs.**
2. **Break Markov Property.**



Testing Accuracy:
90.76%

Interaction Requires Modeling the Human

Explicability: Aim to get π^R closer to π^H (by getting M^R closer to $M^{H)}$



Explanation:

Tell human how to get M^H closer to M^R

--What is the minimum number of changes needed in M^H such that π^R would be optimal plan.

Example 1 – Fetchworld

- Fetch needs to tuck its arms before moving

```
(:action move
        :parameters     (?from ?to - location)
        :precondition   (and (robot-at ?from) (hand-tucked))
        :effect         (and (robot-at ?to) (not (robot-at ?from)))
        )

(:action tuck
        :parameters     ()
        :precondition   (and (not (hand-tucked)))
        :effect         (and (hand-tucked))
        )
```

(move loc2 loc1)
(pickup b1 loc1)
(tuck)
(move loc1 loc2)
(putdown b1 loc2)

Explanation >> **MOVE_LOC1_LOC2-has-precondition-HAND-TUCKED**

```
(:action move
        :parameters     (?from ?to - location)
        :precondition   (and (robot-at ?from))
        :effect         (and (robot-at ?to) (not (robot-at ?from)))
        )
```

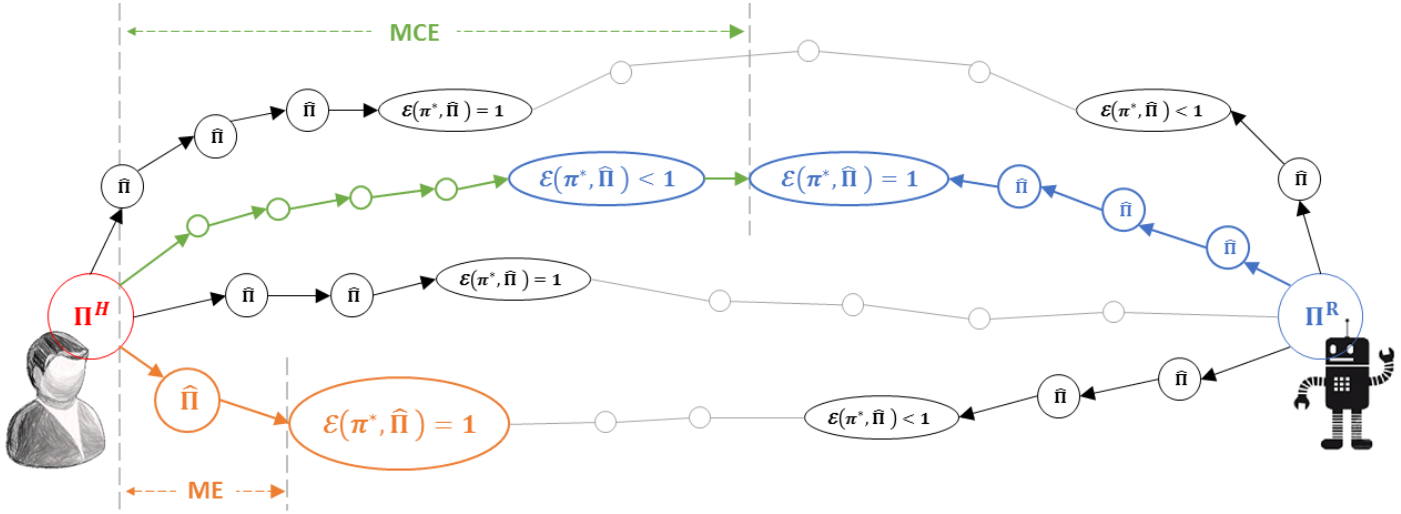
- problem -

```
(:init
        (block-at b1 loc1)
        (robot-at loc2)
        (hand-empty)
        (hand-tucked)
        (charged)
)
```

```
(:goal
(and
        (block-at b1 loc2)
)
)
```

(move loc2 loc1)
(pickup b1 loc1)
(move loc1 loc2)
(putdown b1 loc2)

Minimal Explanation (ME) vs Minimally Complete Explanation (MCE)




```
(:init
(block-at b1 loc1)
(robot-at loc1)
(hand-empty))
```

```
(:goal
(and
(block-at b1 loc2)))
```

plan

```
(pick-up b1)
(tuck)
(move loc1 loc2)
(put-down b1)
```

Robot Model

```
(:action move
:parameters     (?from ?to - location)
:precondition   (and (robot-at ?from)
                     (hand-tucked)
                     (crouched))
:effect         (and (robot-at ?to)
                     (not (robot-at ?from))))
(:action tuck
:parameters     ()
:precondition   ()
:effect         (and (hand-tucked)
                     (crouched)))
(:action crouch
:parameters     ()
:precondition   ()
:effect         (and (crouched)))
```

Human Model

```
(:action move
:parameters     (?from ?to - location)
:precondition   (and (robot-at ?from)
:effect         (and (robot-at ?to)
                     (not (robot-at ?from))))
(:action tuck
:parameters     ()
:precondition   ()
:effect         (and (hand-tucked))

(:action crouch
:parameters     ()
:precondition   ()
:effect         (and (crouched)))
```

ME

```
Explanation >>
MOVE_LOC1_LOC2-has-precondition-HAND-TUCKED
```

MCE

```
Explanation >>
TUCK-has-add-effect-CROUCHED
MOVE_LOC2_LOC1-has-precondition-CROUCHED
```

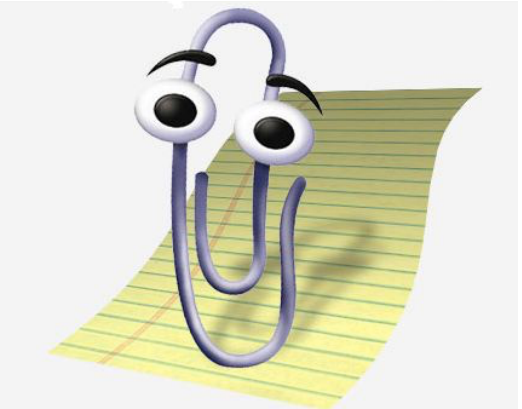
"Beyond Explanations as Soliloquy"
IJCAI 2017

Overview of our ongoing work

- How to learn and plan with incomplete domain models
  - Complete--Approximate--Shallow
- How to plan to be useful to the human
  - Avoiding conflicts and offering serendipitous help
- How to make planned behavior explicable to the human in the loop
  - Humans will parse the behavior in terms of their understanding of the Robot's model
- How to recognize and evaluate what are the desiderata for fluent teaming with humans
  - As the "paper clip" assistant shows, we AI'ers are not great at guessing what humans "like" ☹

Do we really know what (sort of assistance) humans want?

Proactive Help Can be Disconcerting!



Human Factor Studies

- To understand whether human-robot teams perform better with more intelligent/proactive robot teammates or not
- Two studies
  - Wizard-of-Oz Human-Human studies
    - With Cade Bartlett and Nancy Cooke
      - Cade Bartlett's M.S. thesis (in preparation for Journal submission)
  - Human-Planner studies
    - To see if proactive robots that use plan recognition to anticipate human actions help or hinder team performance
      - [IROS 2015][HRI 2015]

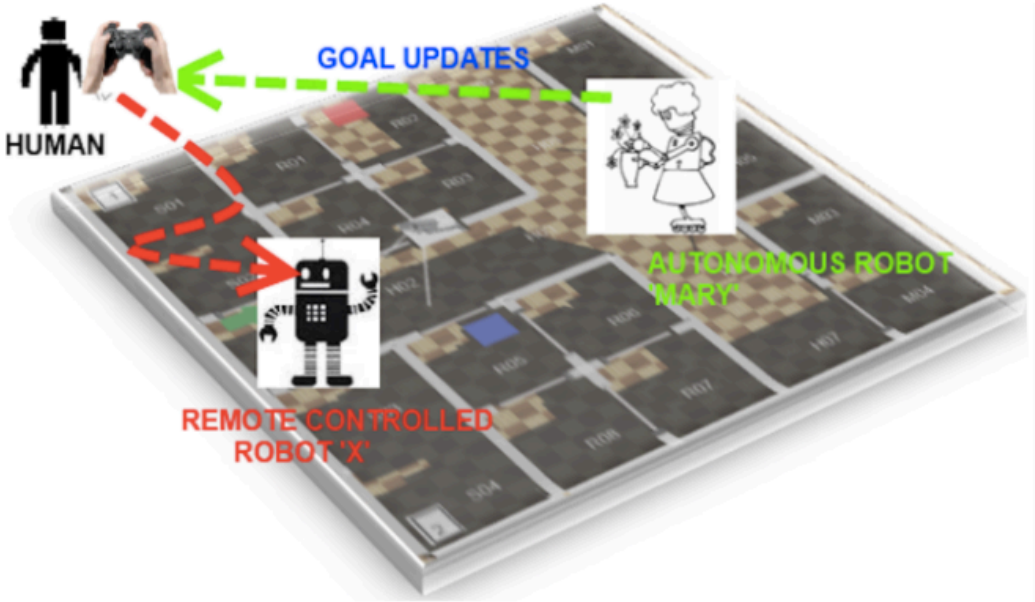
Human-human Teaming Analysis in Urban Search and Rescue

Simulated search task (Minecraft) with human playing role of USAR robot

- 20 internal/external dyads tested
- Conditions of autonomous/intelligent or remotely controlled robot
- Differences in SA, performance, and communications

Analysis of Proactive Support in Human-robot teaming

Simulated search task (Webots) with human remotely controlling a robot while collaborating with an intelligent robot 'Mary':



Findings

Robot with a proactive support capability (vs. without):

- Higher dyad performance
- Lower communication
- Slightly (non-significant) increased mental workload

•Mary with a proactive support capability in our USAR task scenario is generally preferred

[IROS, 2015]

Summary of our ongoing work

- How to learn and plan with incomplete domain models
  - Complete--Approximate--Shallow
- How to plan to be useful to the human
  - Avoiding conflicts and offering serendipitous help
- How to make planned behavior explicable to the human in the loop
  - Humans will parse the behavior in terms of their understanding of the Robot's model
- How to recognize and evaluate what are the desiderata for fluent teaming with humans
  - As the "paper clip" assistant shows, we AI'ers are not great at guessing what humans "like" ☹

 rao@asu.edu  @rao2z  @subbarao2z

Summary

WeChat: Subbarao2z

- Why did AI get so hot now?
  - Progress in perceptual intelligence made AI Technology widely accessible
  - Need to take societal impacts seriously
- Are we done?
  - Commonsense; Incomplete models (and Safety), ability to work with humans..
- Won't somebody please think of the Humans?
  - Human-Aware AI expands the reach and scope of AI
    - Reduces some of the off-the-top worries about AI
    - Brings up novel research challenges
  - Modeling humans in the loop; recognizing their intentions; exhibiting explicable behavior; providing explanations

rao@asu.edu @rao2z @subbarao2z

WeChat: Subbarao2z