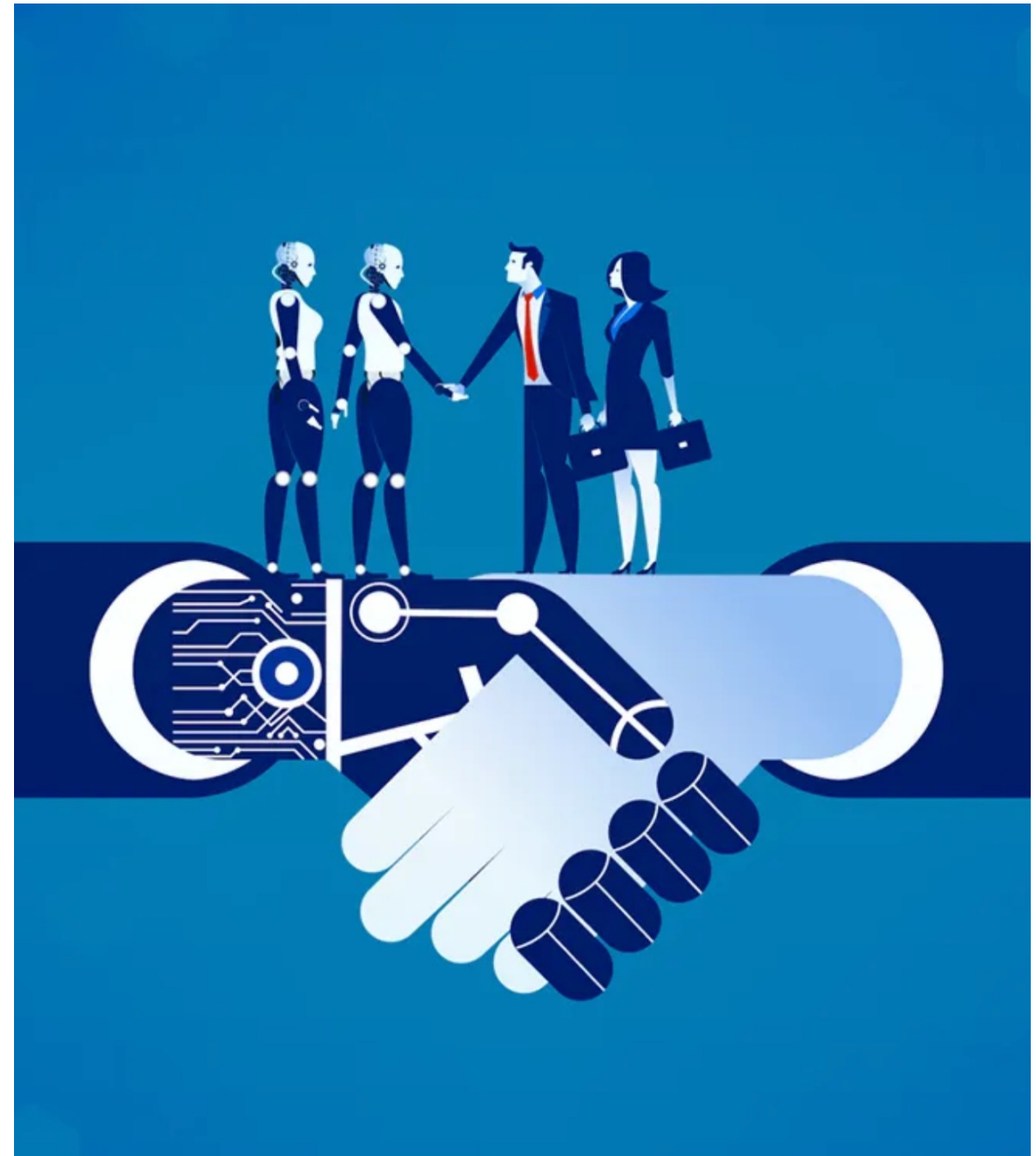


Viewing Autonomic Computing through the Lens of Embodied Artificial Intelligence

Jeff Kephart
Distinguished Researcher & IEEE Fellow
IBM Thomas J Watson Research Center
Yorktown Heights, NY, USA
kephart@us.ibm.com

ACSOS 2023
September 26, 2023



Outline

- Autonomic Computing and Utility Functions
- A fatal flaw
- Embodied AI
- Embodied AI for Autonomic Computing
- Autonomic Computing for Embodied AI
- Conclusions

Outline

- Autonomic Computing and Utility Functions
- A fatal flaw
- Embodied AI
- Embodied AI for Autonomic Computing
- Autonomic Computing for Embodied AI
- Conclusions

Vision of Autonomic Computing

Computing systems that manage themselves in accordance with high-level objectives from humans.

- Self-configuring
- Self-healing
- Self-optimizing
- Self-protecting
- Self-{adapting, organizing, aware, *}

Kephart and Chess, *IEEE Computing*, January 2003.

COVER FEATURE

The Vision of Autonomic Computing

Systems manage themselves according to an administrator's goals. New components integrate as effortlessly as a new cell establishes itself in the human body. These ideas are not science fiction, but elements of the grand challenge to create self-managing computing systems.

Jeffrey O. Kephart
David M. Chess
IBM Thomas J. Watson Research Center

In mid-October 2001, IBM observed that the main progress in the IT industry was complexity crisis.¹ The applications and environments at tens of millions of lines of code required the attention of skilled IT professionals to install and maintain.

The manifesto pointed out that managing today's computing environments beyond the administration of heterogeneous environments into computing systems, and to extend the boundaries into the Internet, is a challenge of complexity. Computing systems appears to be approaching the capability, yet the march toward connectivity and integration rush.

This march could turn the distributed computing—trillions of computers connected to the Internet—into programming language innovations in size and complexity of systems design, but relying solely on current programming methods will not present complexity crisis.

As systems become more diverse, architects are less able to design interactions among components such issues to be dealt with at the most skilled system integrator's level.

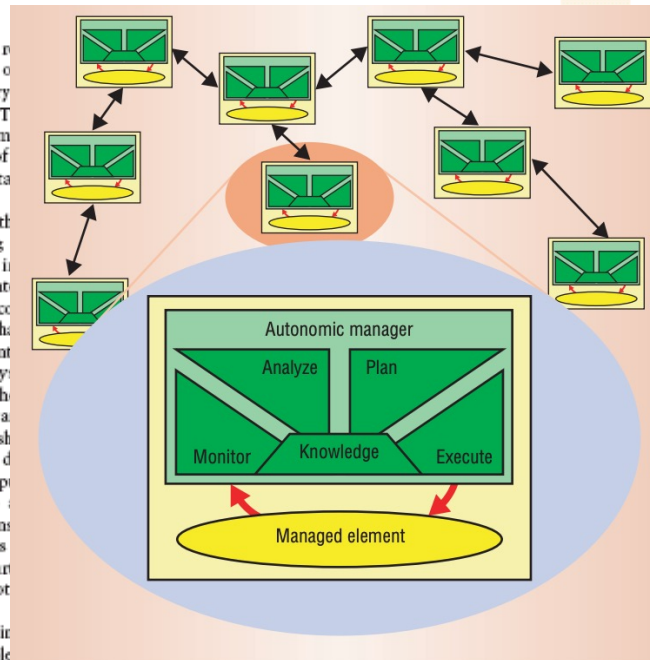


Figure 2. Structure of an autonomic element. Elements interact with other elements and with human programmers via their autonomic managers.

How to *represent* high-level objectives?

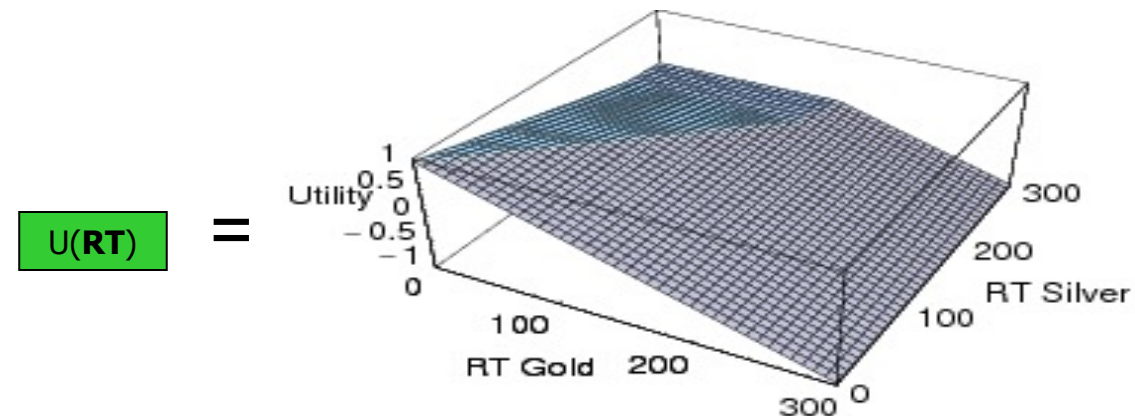
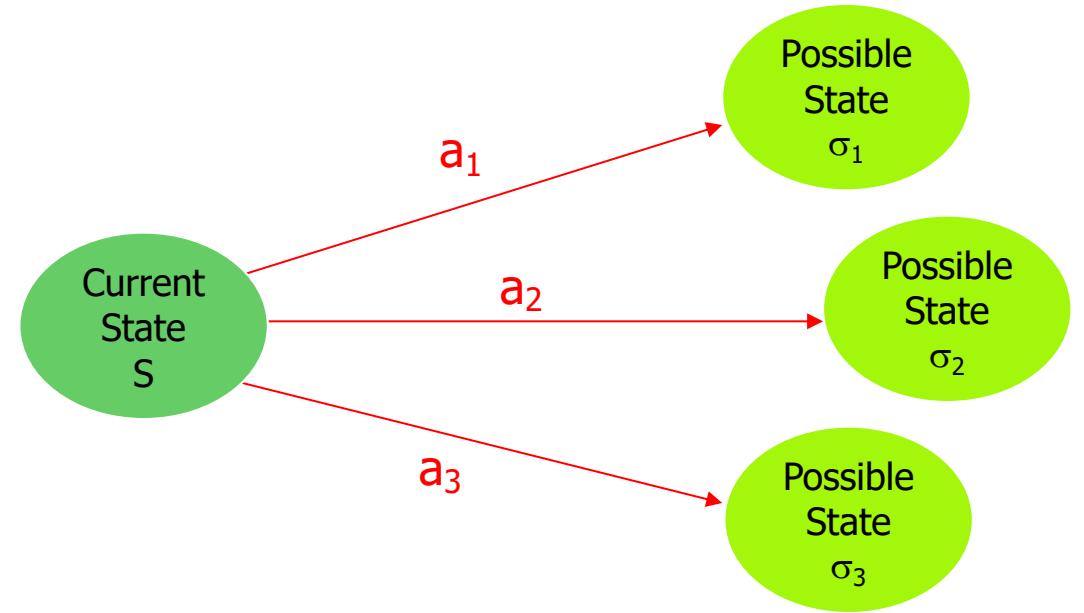
Utility functions map any possible state of a system to a scalar value

They can be obtained from

- Service Level Agreements
- Preference elicitation
- Simple templates

They are a useful representation for high-level objectives

- Value can be transformed and propagated among agents to guide system behavior at multiple levels



Kephart and Walsh, Policy 2004

How to *manage* to high-level objectives?

Elicit utility function $U(S)$ expressed in terms of service attributes S

Model how each attribute S_i depends on controls C and observables O

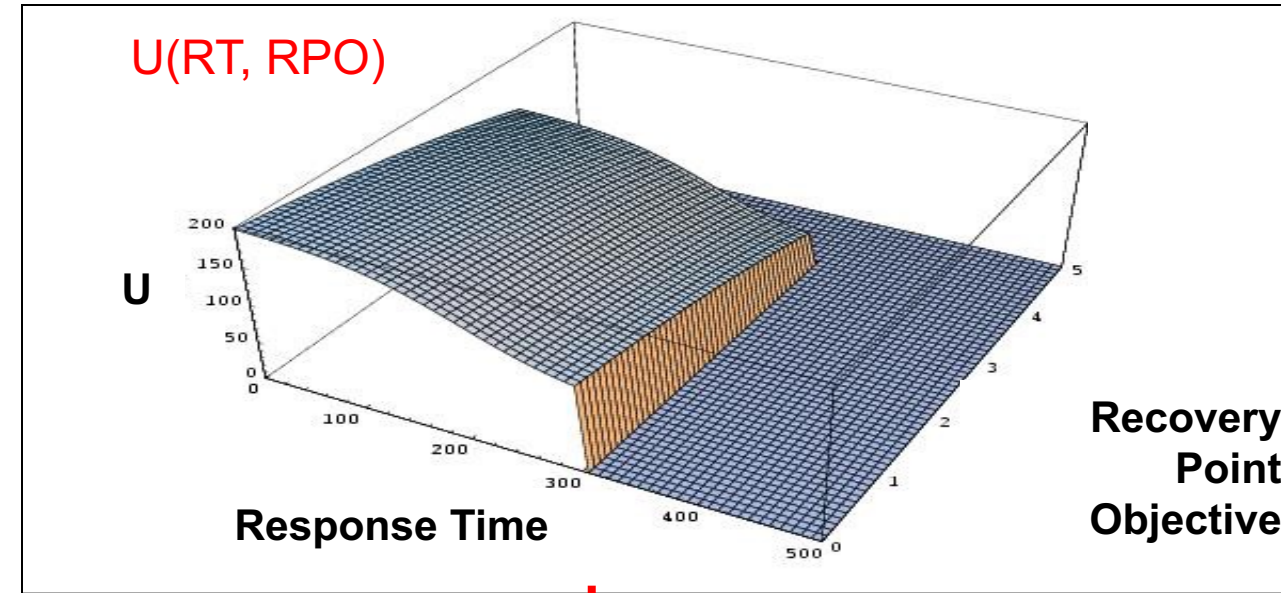
- Models expressed as $S(C; O)$
- E.g., RT(routing weights, request rate)
- Models from experiments, learning, theory

Transform from service utility U to resource utility U' by substitution

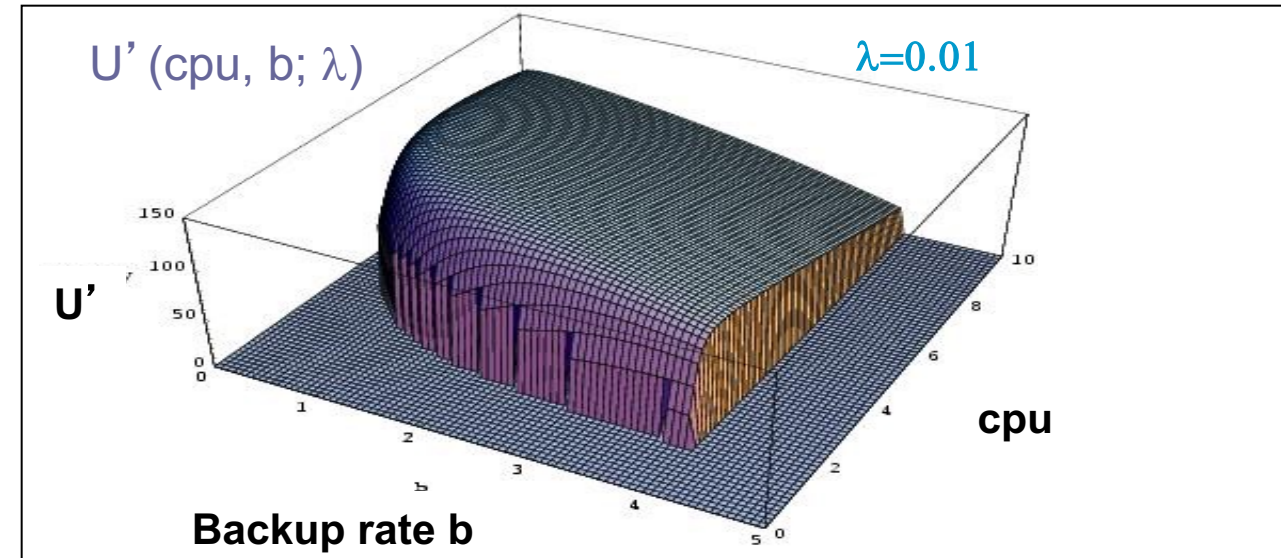
- $U(S) = U(S(C; O)) = U'(C; O)$

Optimize resource utility. As observable O changes, set C to values that maximize $U'(C; O)$

- $C^*(O) = \operatorname{argmax}_C U'(C; O)$
- $U'^*(O) = U'(C^*(O); O)$



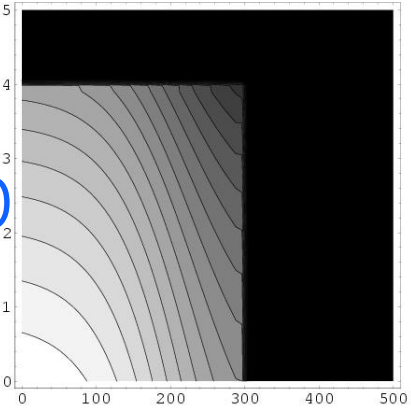
Transform



Finding the optimal control parameters

$U(RT, RPO)$

RPO



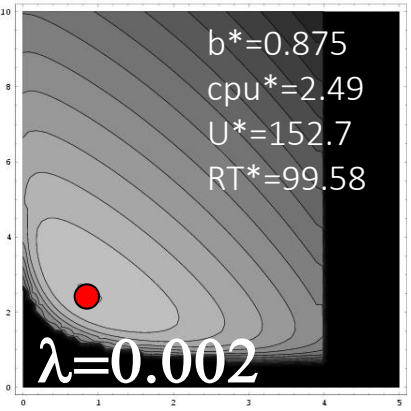
RT

Even if service-level utility remains fixed, resource-level utility depends upon environment.

Thus system responds to environmental changes.

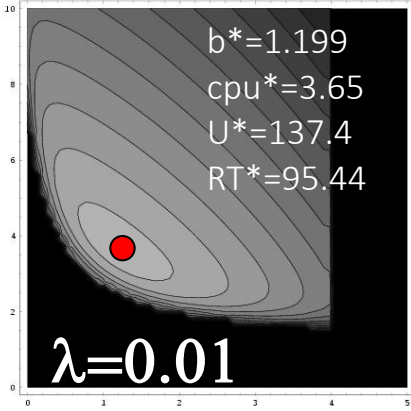
$U'(cpu, b; \lambda)$

cpu



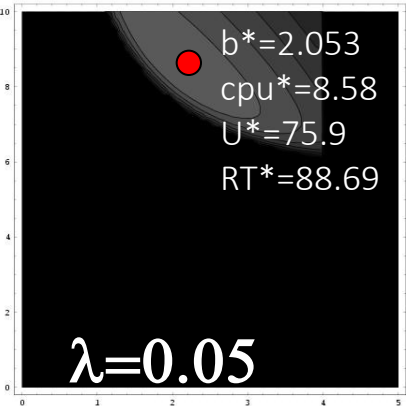
b

cpu



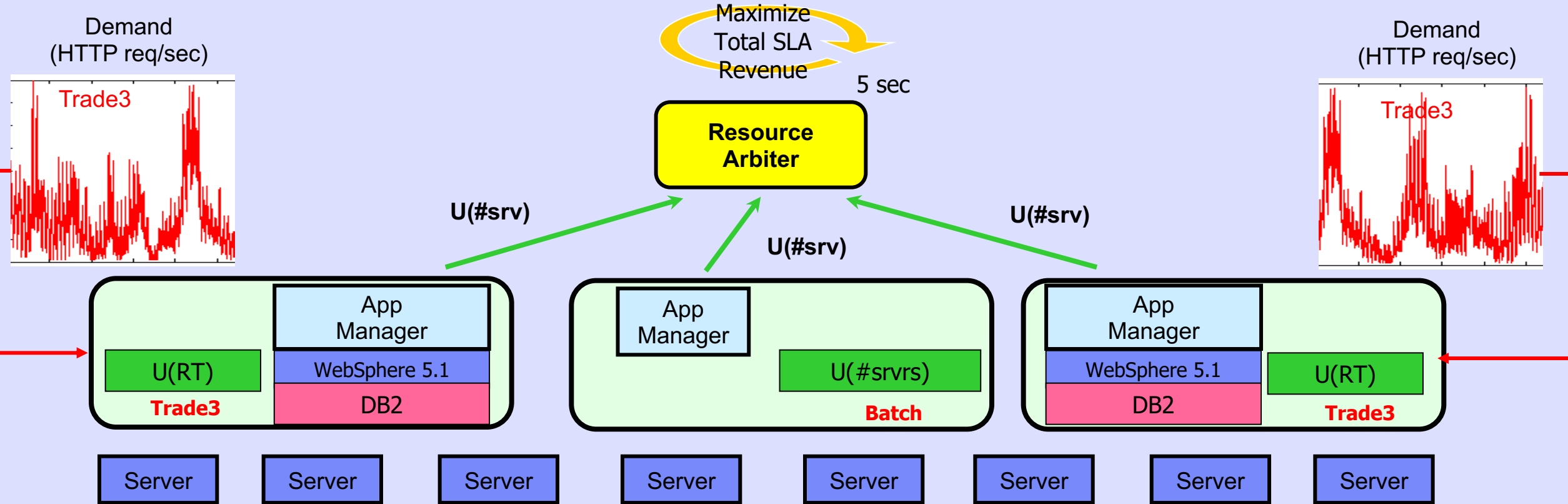
b

cpu



b

Unity Data Center Prototype

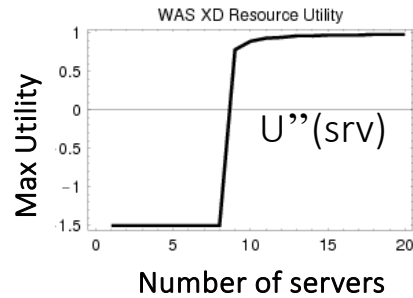


Chess, Segal, Whalley and White, Unity: Experiences with a Prototype Autonomic Computing System, ICAC 2004

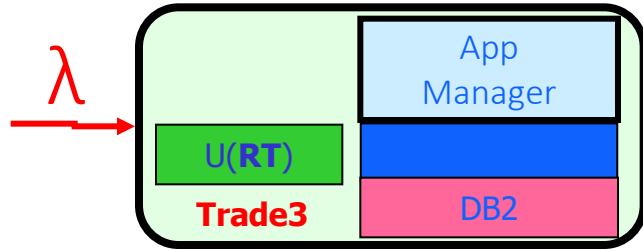
How App Mgr computes its external resource utility

Alternative to generating full curve: utility elicitation

Patrascu, Boutilier et al.
New Approaches to
Optimization and Utility
Elicitation in Autonomic
Computing, AAAI 2005



Resource
Arbiter



Elicit:

$U(\text{RT})$ Service-level utility

My controls Arbiter's controls Observable

Model:

$U(\text{RT}(\mathbf{C}; \text{srv}, \lambda))$

Transform:

$U'(\mathbf{C}; \text{srv}, \lambda) = U(\text{RT}(\mathbf{C}; \text{srv}, \lambda))$

Internal resource-level utility

Optimize:

Optimal internal control settings

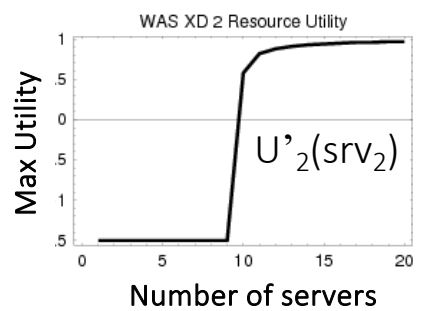
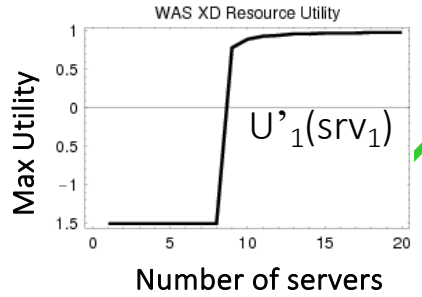
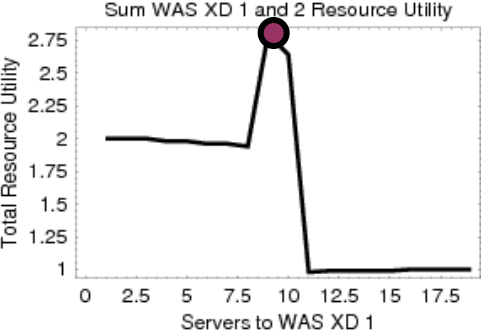
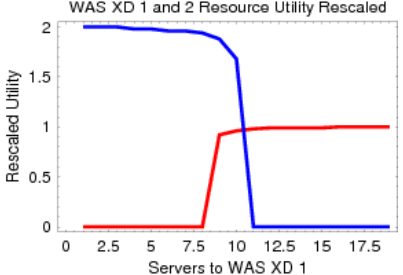
$\mathbf{C}^*(\text{srv}, \lambda) = \text{argmax}_{\mathbf{C}} U'(\mathbf{C}; \text{srv}, \lambda)$

External resource-level utility

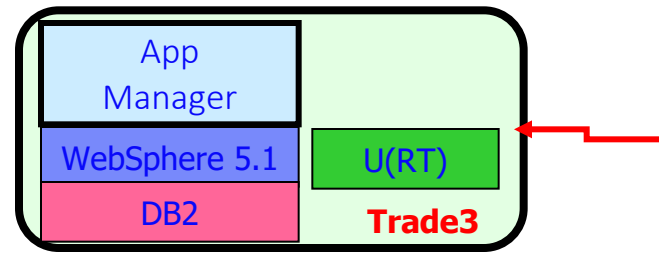
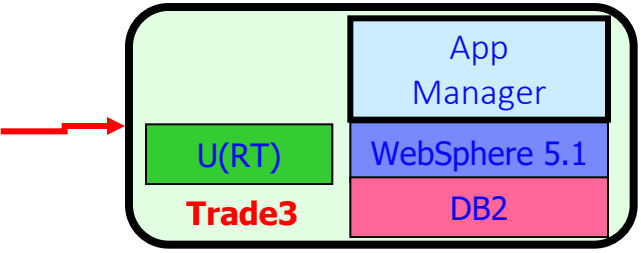
$U''(\text{srv}, \lambda) = U'(\mathbf{C}^*(\text{srv}, \lambda); \text{srv}, \lambda)$

How Arbiter determines final resource allocation

Decision problem:
 Allocate resources
 $srv^* = \operatorname{argmax}_{srv} \sum U''_i(srv_i)$
 Effectively maximizes $\sum U_i(S_i)$



Resource Arbiter



Outline

- Autonomic Computing and Utility Functions
- A fatal flaw
- Embodied AI
- Embodied AI for Autonomic Computing
- Autonomic Computing for Embodied AI
- Conclusions

Athena



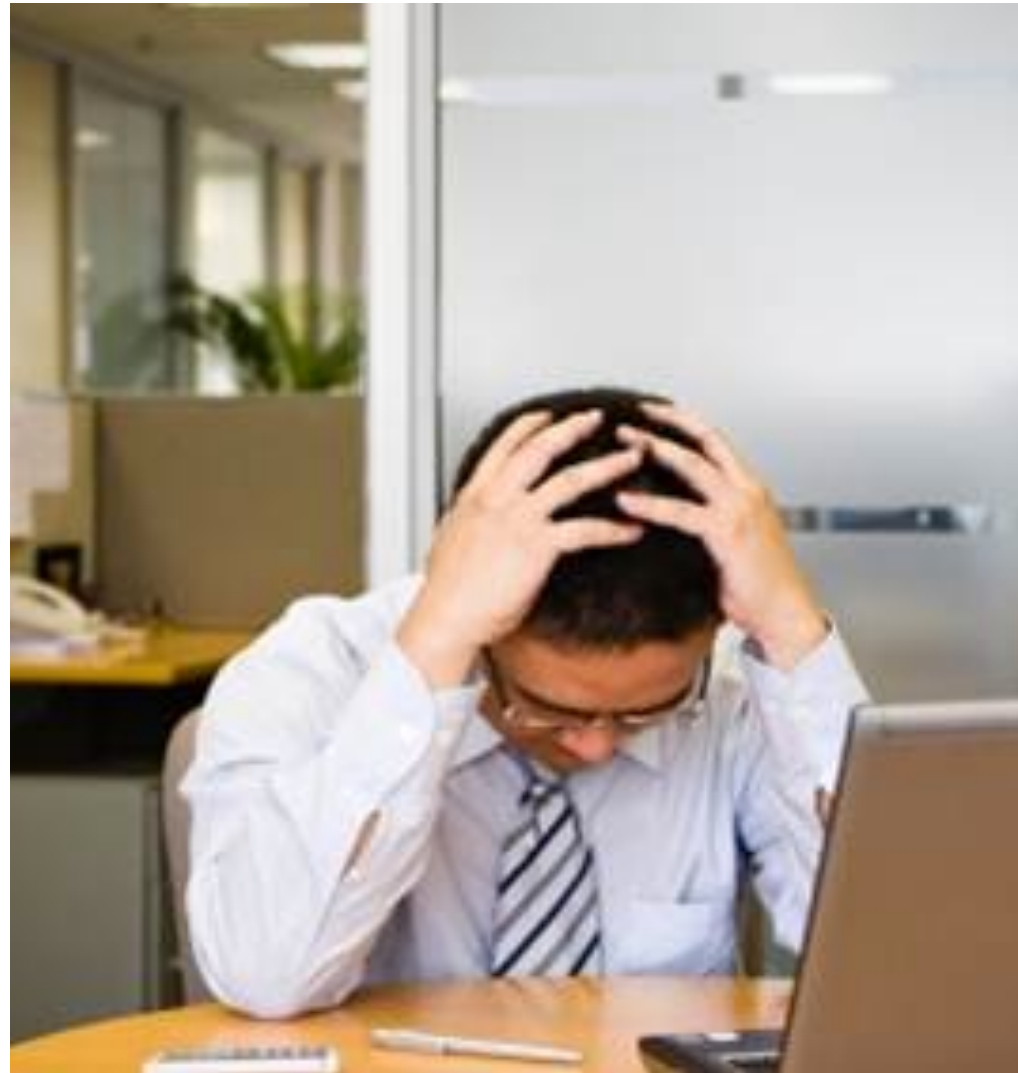
Athena, goddess of wisdom and strategy, emerging fully-formed from head of Zeus

Athena



Athena, goddess of wisdom and strategy, emerging fully-formed from head of Zeus

System Admin



Utility function, the locus of wisdom and strategy, emerging fully-formed from head of modern sys admin

Athena



Athena, goddess of wisdom and strategy, emerging fully-formed from head of Zeus

System A



Utility function, the locus of wisdom and strategy, emerging fully-formed from head of modern sys admin

Reality Check: Humans as Decision Makers



The Science of Irrationality

- Tversky and Kahneman [1974]
 - well-informed and intelligent humans are subject to surprising cognitive biases and gaps in rationality that can lead to incorrect inferences and suboptimal decisions
- Ariely, *Predictably Irrational* [2008]
- Dozens of types of cognitive bias have been cataloged
 - List of recognized cognitive biases

Nobel Prize in Economics, 2002

Tversky and Kahneman Science, 1974

Judgment under Uncertainty: Heuristics and Biases

Biases in judgments reveal some heuristics of thinking under uncertainty.

Amos Tversky and Daniel Kahneman

Many decisions are based on beliefs concerning the likelihood of uncertain events such as the outcome of an election, the guilt of a defendant, or the future value of the dollar. These beliefs are usually expressed in statements such as "I think that . . ." "chances are . . .," "it is unlikely that . . .," and so forth. Occasionally, beliefs concerning uncertain events are expressed in numerical form as odds or subjective probabilities. What determines such beliefs? How do people assess the probability of an uncertain event or the value of an uncertain quantity? This article shows that people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors.

Many decisions are based on beliefs concerning the likelihood of uncertain events such as the outcome of an election, the guilt of a defendant, or the future value of the dollar. These beliefs are usually expressed in statements such as "I think that . . ." "chances are . . .," "it is unlikely that . . .," and so forth. Occasionally, beliefs concerning uncertain events are expressed in numerical form as odds or subjective probabilities. What determines such beliefs? How do people assess the probability of an uncertain event or the value of an uncertain quantity? This article shows that people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors.

Representativeness

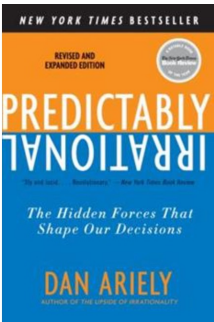
Many of the probabilistic judgments with which people are concerned are based on one of the following types of heuristics: the probability that object A belongs to class B? What is the probability that event A originates from p . . . What is the probability that . . .

occupation from a list of possibilities (for example, farmer, salesman, airline pilot, librarian, or physician)? How do people order these occupations from most to least likely? In the representativeness heuristic, the probability that Steve is a librarian, for example, is assessed by the degree to which he is representative of, or similar to, the stereotype of a librarian. Indeed, research with problems of this type has shown that people order the occupations by probability and by similarity in exactly the same way (1). This approach to the judgment of probability leads to serious errors, because sim-



Cognitive Bias example

(Experiment conducted on two groups of 100 MIT Sloan students)

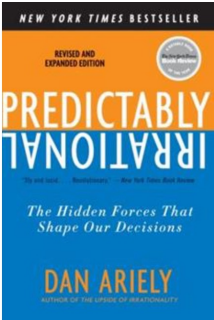


Dan Ariely,
Predictably Irrational (Harper Collins 2008)

Economist.com	SUBSCRIPTIONS
OPINION	Welcome to The Economist Subscription Centre
WORLD	Pick the type of subscription you want to buy or renew.
BUSINESS	<input type="checkbox"/> Economist.com subscription - US \$59.00
FINANCE & ECONOMICS	One-year subscription to Economist.com.
SCIENCE & TECHNOLOGY	Includes online access to all articles from <i>The Economist</i> since 1997.
PEOPLE	<input type="checkbox"/> Print subscription - US \$125.00
BOOKS & ARTS	One-year subscription to the print edition of <i>The Economist</i> .
MARKETS & DATA	<input type="checkbox"/> Print & web subscription - US \$125.00
DIVERSIONS	One-year subscription to the print edition of <i>The Economist</i> and online access to all articles from <i>The Economist</i> since 1997.

Cognitive Bias example

(Experiment conducted on two groups of 100 MIT Sloan students)



Dan Ariely,
Predictably Irrational (Harper Collins 2008)

Economist.com	
OPINION	<h2>SUBSCRIPTIONS</h2> <p>Welcome to The Economist Subscription Centre</p> <p>Pick the type of subscription you want to buy or renew.</p> <ul style="list-style-type: none"><input type="checkbox"/> Economist.com subscription - US \$59.00 One-year subscription to Economist.com. Includes online access to all articles from <i>The Economist</i> since 1997.<input type="checkbox"/> Print subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i>.<input type="checkbox"/> Print & web subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i> and online access to all articles from <i>The Economist</i> since 1997.
WORLD	
BUSINESS	
FINANCE & ECONOMICS	
SCIENCE & TECHNOLOGY	
PEOPLE	
MARKETS	
DATA	
DIVERSIONS	

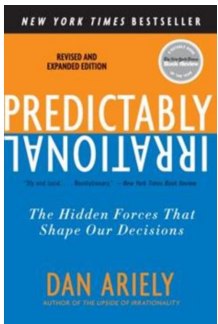
16

0

84

Cognitive Bias example

(Experiment conducted on two groups of 100 MIT Sloan students)



Dan Ariely,
Predictably Irrational (Harper Collins 2008)

Economist.com	SUBSCRIPTIONS
OPINION	<p>Welcome to The Economist Subscription Centre</p> <p>Pick the type of subscription you want to buy or renew.</p> <p><input type="checkbox"/> Economist.com subscription - US \$59.00 One-year subscription to Economist.com. Includes online access to all articles from <i>The Economist</i> since 1997.</p> <p><input type="checkbox"/> Print subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i>.</p> <p><input type="checkbox"/> Print & web subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i> and online access to all articles from <i>The Economist</i> since 1997.</p>
WORLD	
BUSINESS	
FINANCE & ECONOMICS	
SCIENCE & TECHNOLOGY	
PEOPLE	
BOOKS & ARTS	
MARKETS & DATA	
DIVERSIONS	

16

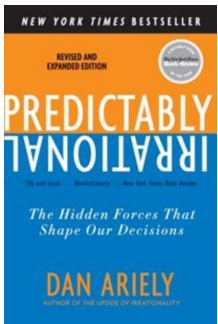
0

84

Economist.com	SUBSCRIPTIONS
OPINION	<p>Welcome to The Economist Subscription Centre</p> <p>Pick the type of subscription you want to buy or renew.</p> <p><input type="checkbox"/> Economist.com subscription - US \$59.00 One-year subscription to Economist.com. Includes online access to all articles from <i>The Economist</i> since 1997.</p> <p><input type="checkbox"/> Print & web subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i> and online access to all articles from <i>The Economist</i> since 1997.</p>
WORLD	
BUSINESS	
FINANCE & ECONOMICS	
SCIENCE & TECHNOLOGY	
PEOPLE	
BOOKS & ARTS	
MARKETS & DATA	
DIVERSIONS	

Cognitive Bias example

(Experiment conducted on two groups of 100 MIT Sloan students)



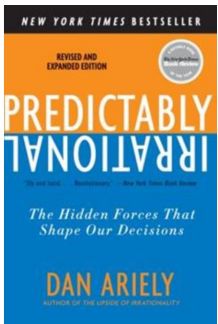
Dan Ariely,
Predictably Irrational (Harper Collins 2008)

Economist.com	
SUBSCRIPTIONS	
OPINION	<p>Welcome to The Economist Subscription Centre</p> <p>Pick the type of subscription you want to buy or renew.</p> <p><input type="checkbox"/> Economist.com subscription - US \$59.00 One-year subscription to Economist.com. Includes online access to all articles from <i>The Economist</i> since 1997.</p> <p><input type="checkbox"/> Print subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i>.</p> <p><input type="checkbox"/> Print & web subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i> and online access to all articles from <i>The Economist</i> since 1997.</p>
WORLD	
BUSINESS	
FINANCE & ECONOMICS	
SCIENCE & TECHNOLOGY	
OPINION	
WORLD	
BUSINESS	
FINANCE & ECONOMICS	
SCIENCE & TECHNOLOGY	
16	
0	
84	

Economist.com	
SUBSCRIPTIONS	
OPINION	<p>Welcome to The Economist Subscription Centre</p> <p>Pick the type of subscription you want to buy or renew.</p> <p><input type="checkbox"/> Economist.com subscription - US \$59.00 One-year subscription to Economist.com. Includes online access to all articles from <i>The Economist</i> since 1997.</p> <p><input type="checkbox"/> Print & web subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i> and online access to all articles from <i>The Economist</i> since 1997.</p>
WORLD	
BUSINESS	
FINANCE & ECONOMICS	
SCIENCE & TECHNOLOGY	
OPINION	
WORLD	
BUSINESS	
FINANCE & ECONOMICS	
SCIENCE & TECHNOLOGY	
68	
32	

Cognitive Bias example

(Experiment conducted on two groups of 100 MIT Sloan students)



Dan Ariely,
Predictably Irrational (Harper Collins 2008)

Decoy effect

Economist.com	
SUBSCRIPTIONS	
OPINION	<p>Welcome to The Economist Subscription Centre</p> <p>Pick the type of subscription you want to buy or renew.</p> <p><input type="checkbox"/> Economist.com subscription - US \$59.00 One-year subscription to Economist.com. Includes online access to all articles from <i>The Economist</i> since 1997.</p> <p><input type="checkbox"/> Print subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i>.</p> <p><input type="checkbox"/> Print & web subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i> and online access to all articles from <i>The Economist</i> since 1997.</p>
WORLD	
BUSINESS	
FINANCE & ECONOMICS	
SCIENCE & TECHNOLOGY	
PEOPLE	
SPORTS	
MARKETS & DATA	
DIVERSIONS	

16

0

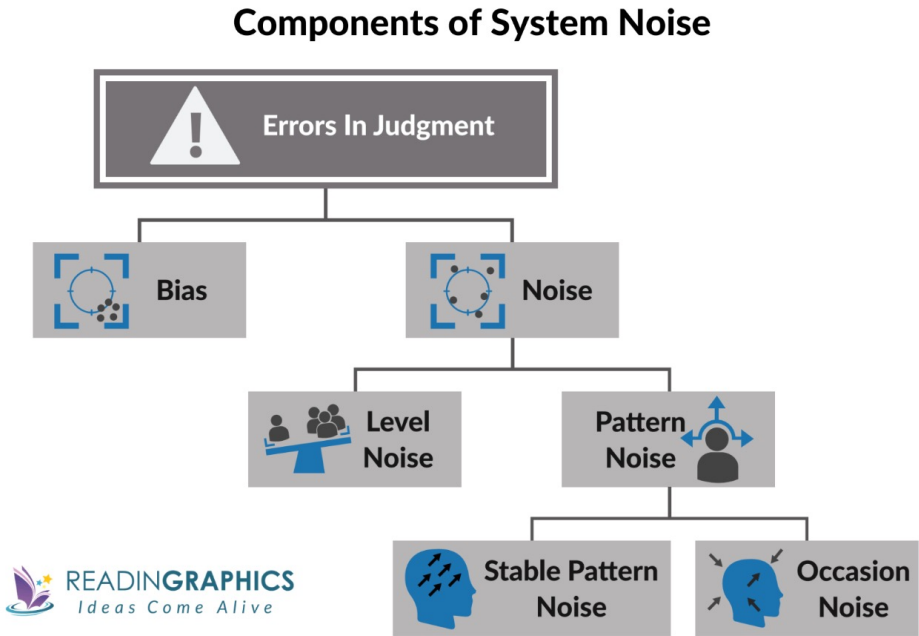
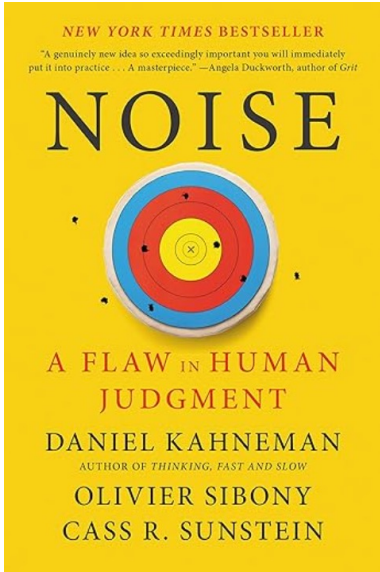
84

Economist.com	
SUBSCRIPTIONS	
OPINION	<p>Welcome to The Economist Subscription Centre</p> <p>Pick the type of subscription you want to buy or renew.</p> <p><input type="checkbox"/> Economist.com subscription - US \$59.00 One-year subscription to Economist.com. Includes online access to all articles from <i>The Economist</i> since 1997.</p> <p><input type="checkbox"/> Print & web subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i> and online access to all articles from <i>The Economist</i> since 1997.</p>
WORLD	
BUSINESS	
FINANCE & ECONOMICS	
SCIENCE & TECHNOLOGY	
PEOPLE	
SPORTS	
MARKETS & DATA	
DIVERSIONS	

68

32

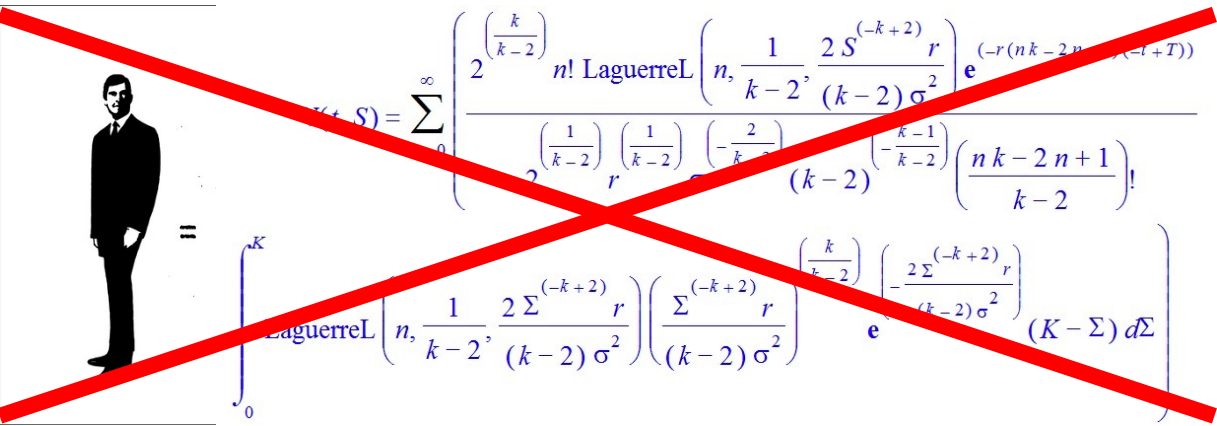
Noise (Kahneman et al. 2022)



- Human judgments are highly inconsistent across and within individuals
- Human decision makers are highly overconfident

High-level objectives from humans, reconsidered

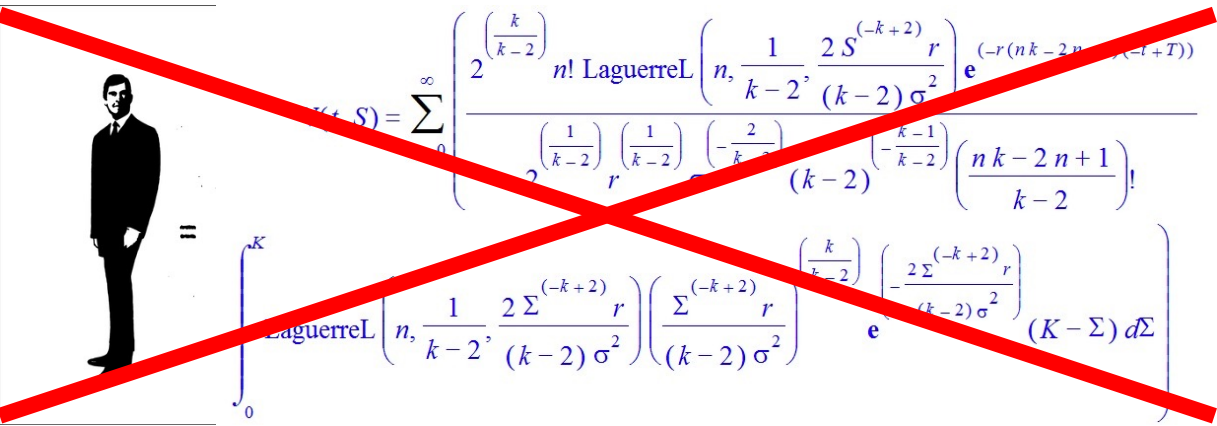
We aren't very good at making decisions that optimize our objectives



Homo economicus

High-level objectives from humans, reconsidered

We aren't very good at making decisions that optimize our objectives



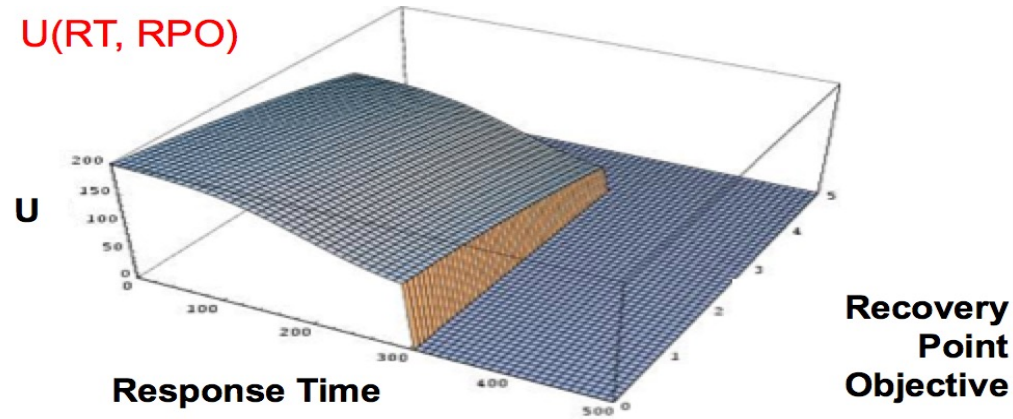
Homo economicus



Homo not-so-sapiens

High-level objectives from humans, reconsidered

We aren't very good at expressing our objectives (utilities) in a mathematical form



High-level objectives from humans, reconsidered

Often, we aren't even sure what our objectives are

Greg Ganger (CMU) introduced notion of *complaint-based tuning*:

“Humans are not good at precisely specifying what they want, but they are very good at complaining when they are not satisfied.”

[Ganger & Strunk, AASMS 2003]

Economist.com	
OPINION	SUBSCRIPTIONS Welcome to The Economist Subscription Centre Pick the type of subscription you want to buy or renew. <input type="checkbox"/> Economist.com subscription - US \$59.00 One-year subscription to Economist.com. Includes online access to all articles from <i>The Economist</i> since 1997. <input type="checkbox"/> Print subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i> . <input type="checkbox"/> Print & web subscription - US \$125.00 One-year subscription to the print edition of <i>The Economist</i> and online access to all articles from <i>The Economist</i> since 1997.
WORLD	
BUSINESS	
FINANCE & ECONOMICS	
SCIENCE & TECHNOLOGY	
PEOPLE	
MARKETS	
MARKETING DATA	
DIVERSIONS	

16

0

84

High-level objectives from humans, reconsidered

If humans can't explicitly specify their high-level objectives, is the vision of utility-based AC dead?

IBM Research

How to represent high-level policies? Kephart and Walsh, Policy04

- Utility function maps any possible state of a system to a scalar value
- They can be obtained from
 - Service Level Agreement
 - preference elicitation
 - simple templates
- They are a very useful representation for high-level objectives
 - Value can be transformed and aggregated among agents to guide system behavior

$U(\mathbf{RT}) =$

High-level objectives from humans, reconsidered

If humans can't explicitly specify their high-level objectives, is the vision of utility-based AC dead?

IBM Research

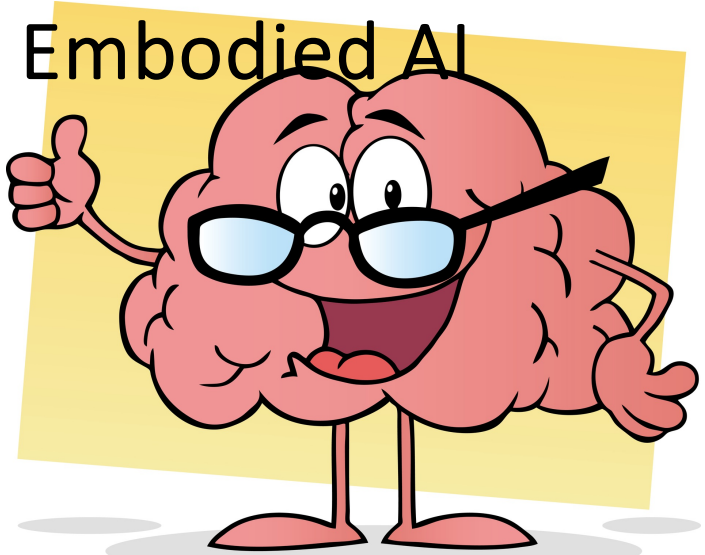
How to represent high-level policies? Kephart and Walsh, Policy04

- Utility functions assign a numerical value to any possible state of a system
 - Service Level Agreement
 - preference elicitation
 - simple templates
- They can be obtained from
 - Value can be transformed and aggregated among agents to guide system behavior
- They are a very useful representation for high-level objectives

$U(RT) =$

No.

But we must reconsider how humans and autonomic systems should interact...



Outline

- Autonomic Computing and Utility Functions
- A fatal flaw
- **Embodied AI**
- Embodied AI for Autonomic Computing
- Autonomic Computing for Embodied AI
- Conclusions

Embodied AI

- Vision
- Anatomy
- Prototypes
- Research challenges
 - Leveraging Large Language Models
 - New non-verbal modalities

Embodied AI: Multi-modal AI Assistants



Image source: <https://www.concentrix.com/blog/workplace-diversity-road-equality-long-good-start/>

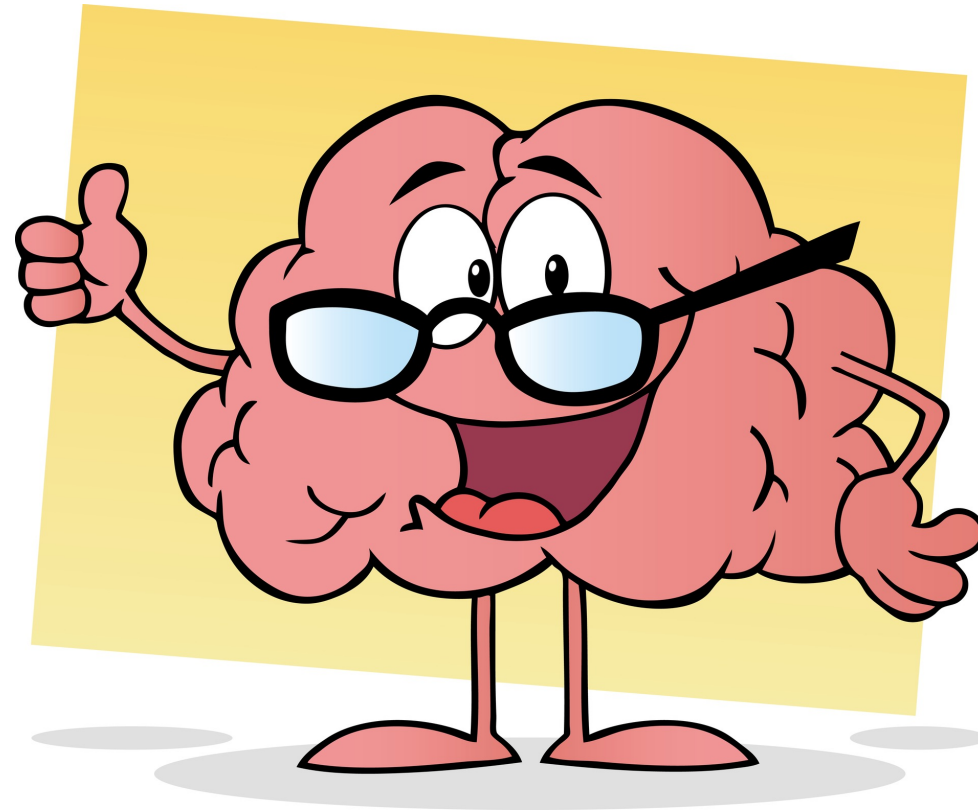
- Human-human communication is multi-modal
 - Speech
 - Pointing
 - Gesture
 - Head orientation
 - Eye contact
 - Facial expression

To collaborate effectively with humans, AI agents need to communicate multi-modally

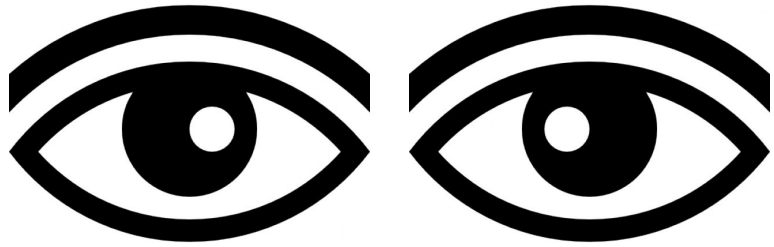
Embodied AI

- Vision
- **Anatomy**
- Prototypes
- Research challenges
 - Leveraging Large Language Models
 - New non-verbal modalities

Anatomy of an Embodied Agent



Anatomy of an Embodied Agent



Sensors



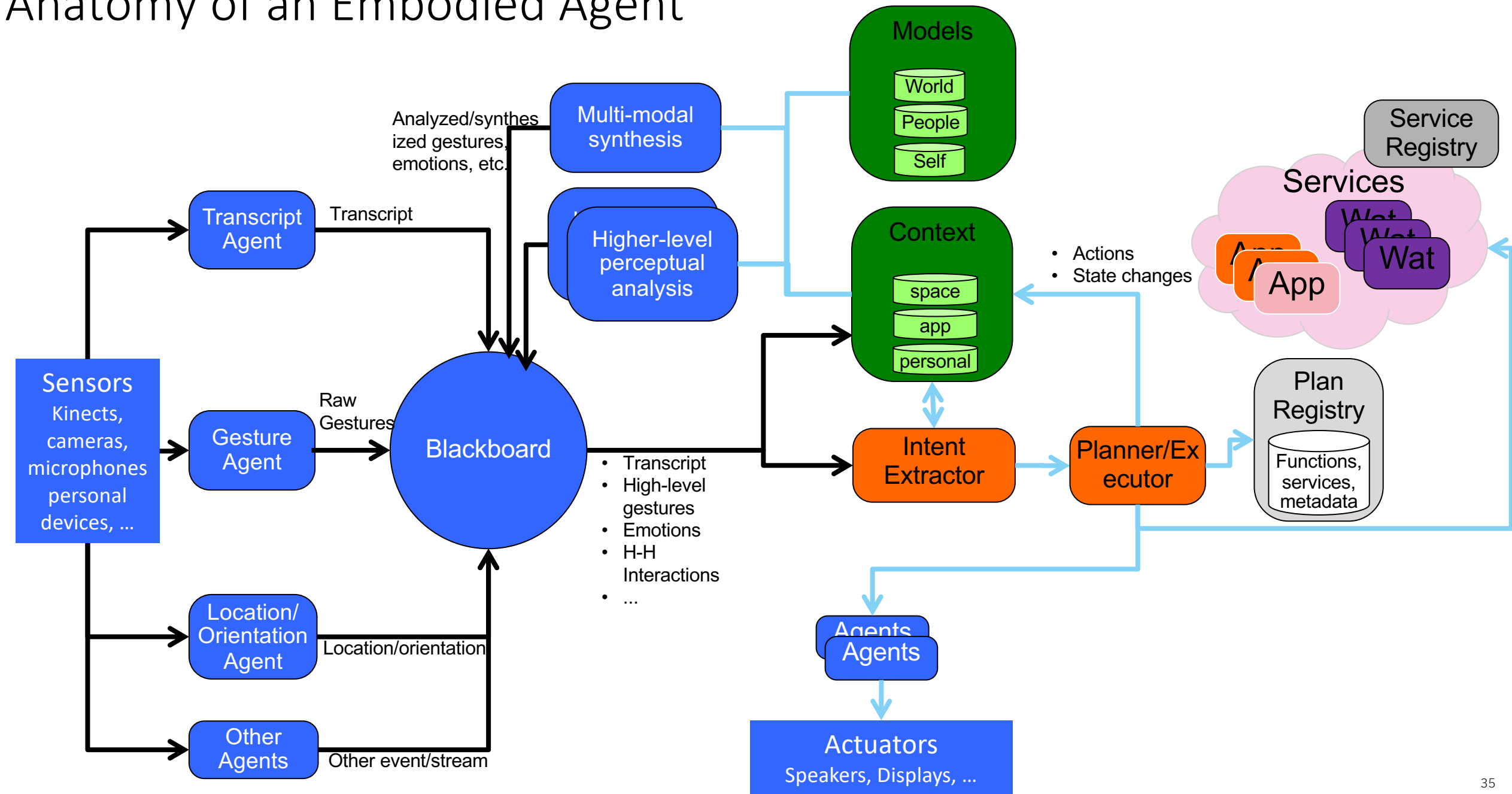
Intelligence



Effectors



Anatomy of an Embodied Agent

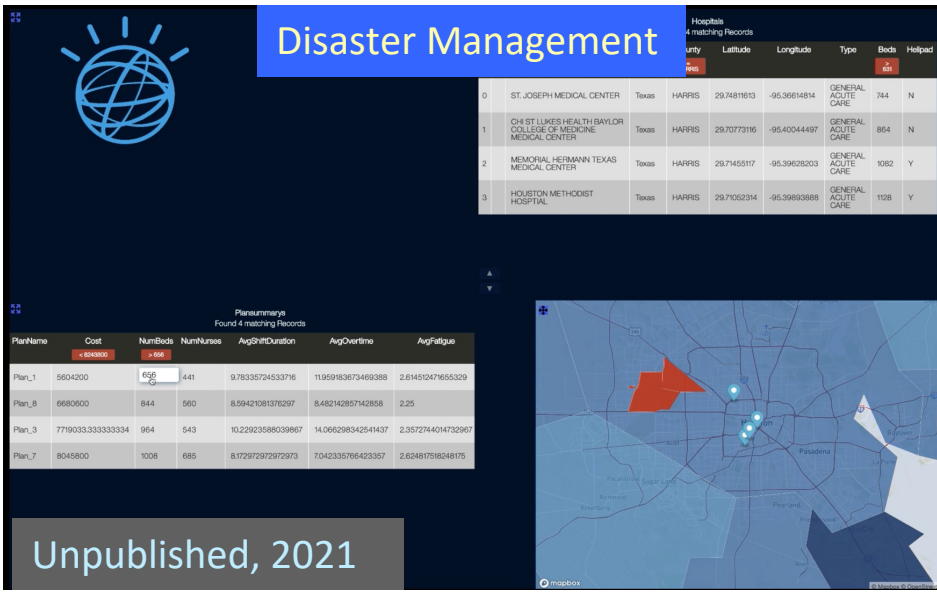


Embodied AI

- Vision
- Anatomy
- **Prototypes**
- Research challenges
 - Leveraging Large Language Models
 - New non-verbal modalities

Multi-modal Assistant prototypes

Kephart, Jeffrey O. "Multi-modal agents for business intelligence." *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*. 2021.



[video]

Cognitive Virtual Operations Assistant

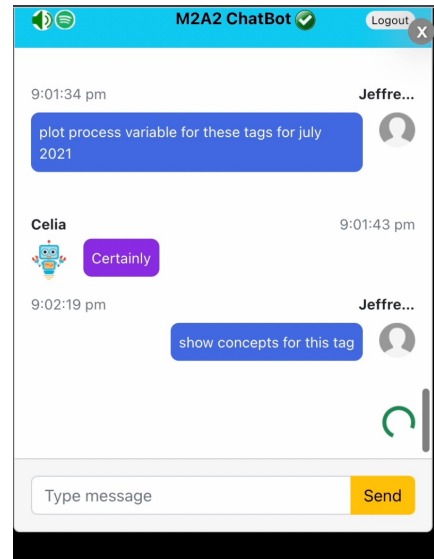
The cognitive VOA provides operators with a coherent and contextual view of multiple data sources

- Time series
- Alarm data
- DCS structure
- P&ID diagrams
- Manuals
- ...

The VOA helps operators

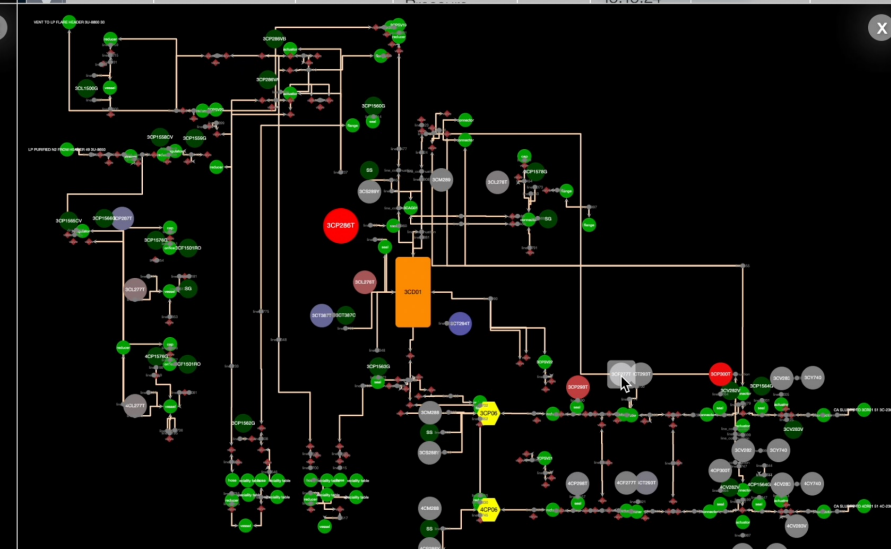
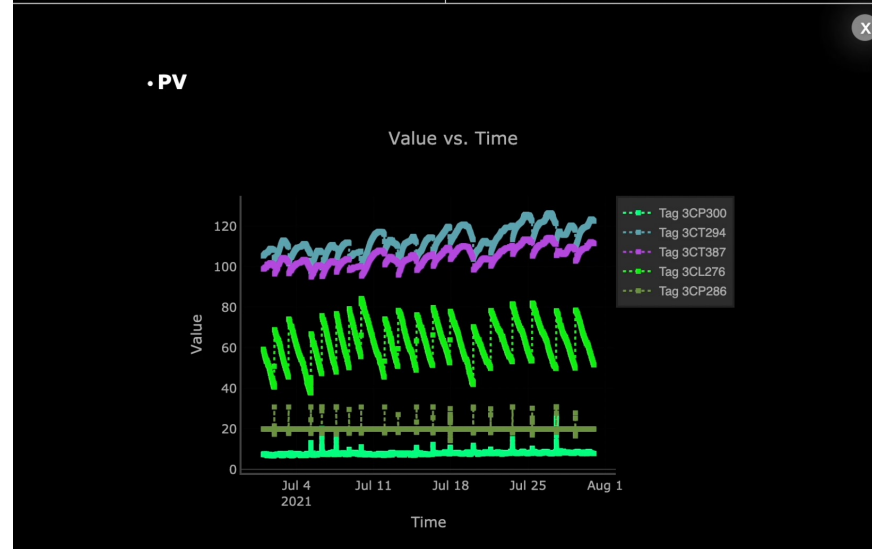
- monitor system behavior
- discover correlations and trends
- diagnose and anticipate problems

Operators interact with the VOA through natural language (speech, text and pointing)



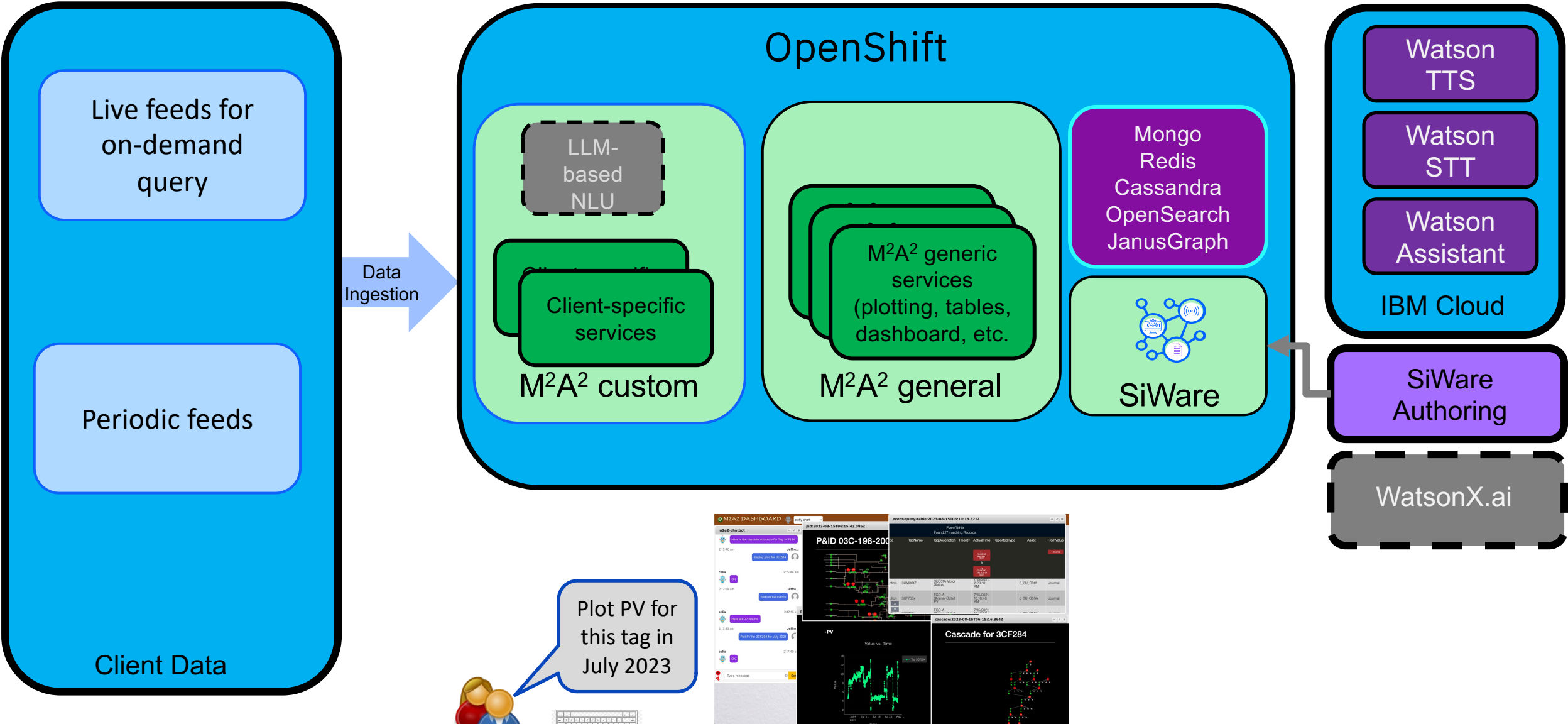
The screenshot shows an "Event Table" with a search filter for "pressure". It displays a list of events with columns: Id, EventType, TagName, TagDescription, Priority, ActualTime, ReportedType, and Asset. The table contains several rows of data, including alarms and acknowledgments for tags like 3AP211 and 3AP247.

Id	EventType	TagName	TagDescription	Priority	ActualTime	ReportedType	Asset
0	Alarm	3AP211	COMP B Suction Pressure	High	7/17/2021, 15:16:07	PVHIGH	1_3A_C01
1	Acknowledge	3AP211	COMP B Suction Pressure		7/17/2021, 15:16:16	PVHIGH	1_3A_C01
2	ReturnToNormal	3AP211	COMP B Suction Pressure	High	7/17/2021, 15:17:31	PVHIGH	1_3A_C01
3	Alarm	3AP247	COMP A Suction Pressure	High	7/17/2021, 15:16:24	PVHIGH	1_3A_C01

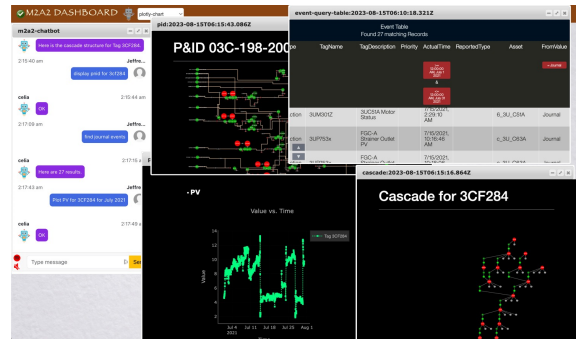


The prototype is being developed by IBM Research in collaboration with multiple industrial partners

VOA Data Flow and Architecture



Plot PV for this tag in July 2023



Thin client

The multi-modal advantage

Text/speech only

Id	Date	Sensor	AlarmType	Location	Description	
6	1455767	2/28/2022, 21:01:20	3YK794	TooLow	Buffalo	LN2 flow; Plant B10 Floor 2 Rm 203
7	1455513	2/28/2022, 20:00:03	3HO790W52			LN2 flow; Plant A2 Floor 1 Rm 101
8	1455492	2/28/2022, 20:00:01	3HO790W52			LN2 flow; Plant A2 Floor 1 Rm 101
9	1455477	2/28/2022, 19:59:21	3PD870Z	TooHigh	Troy	LN2 flow; Plant A2 Floor 1 Rm 101

Show incidents with date in the range February 1 2022 to February 28 2022 pertaining to LN2 for which AlarmType is TooHigh.

Text: 122 characters

Speech: 11.8 sec

Cog. burden: Medium

Text/speech + pointing + context

Id	Date	Sensor	AlarmType	Location	Description	
7	1455513	2/28/2022, 20:00:03	3HO790W52			LN2 flow; Plant A2 Floor 1 Rm 101
8	1455492	2/28/2022, 20:00:01	3HO790W52			LN2 flow; Plant A2 Floor 1 Rm 101
9	1455477	2/28/2022, 19:59:21	3PD870Z	TooHigh	Troy	LN2 flow; Plant A2 Floor 1 Rm 101
10	1455468	2/28/2022, 19:57:48	3PD870Z		Troy	LN2 flow; Plant A2 Floor 1 Rm 101

Text: 24 characters

Speech: 1.8 sec

Cog. burden: Low

Show incidents like this.

Embodied AI

- Vision
- Anatomy
- Prototypes
- **Research challenges**
 - **Leveraging Large Language Models**
 - **New non-verbal modalities**

Research Challenge: Leveraging Large Language Models



Write a brief sentence explaining what is ChatGPT



ChatGPT is an AI-powered language model developed by OpenAI, designed to generate human-like text and engage in natural language conversations on a wide range of topics.

Leveraging Large Language Models

- **ChatGPT 3.5:** Self-supervised training on an Internet-scale corpus yields a model that does a very impressive job of answering general knowledge questions
- But it (and all LLMs) often “hallucinate” answers
- Industry applications require more refinements
 - Fine-tuning
 - Prompt engineering
 - Retrieval-augmented generation

JE

Write a short introduction to a talk on autonomic computing and embodied AI assistants



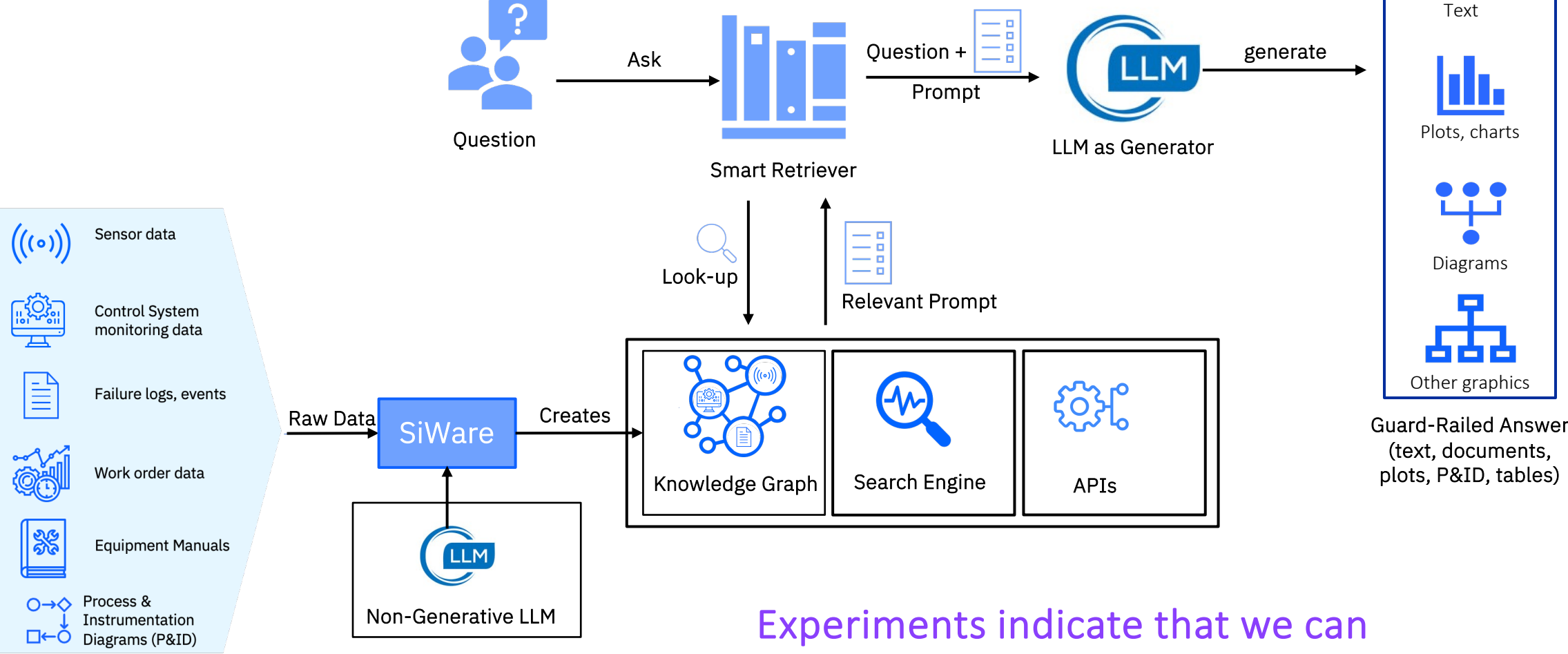
Ladies and gentlemen, esteemed colleagues, and technology enthusiasts, welcome to today's talk on the fascinating intersection of autonomic computing and embodied AI assistants. In an era where technology continues to evolve at an unprecedented pace, we find ourselves on the cusp of a transformative revolution in the way we interact with machines and the level of autonomy we can bestow upon them.

Autonomic computing, a concept inspired by the human autonomic nervous system, seeks to create self-managing systems capable of adapting, optimizing, and healing themselves without constant human intervention. On the other hand, embodied AI assistants represent the pinnacle of human-machine collaboration, integrating AI into physical embodiments to facilitate more natural and intuitive interactions.

In this talk, we will delve into the symbiotic relationship between these two domains, exploring how autonomic computing principles can empower embodied AI assistants to perform increasingly complex tasks, adapt to dynamic environments, and even anticipate our needs. We will discuss the potential applications, ethical considerations, and the exciting possibilities that lie ahead as we merge the worlds of autonomy and embodiment in AI.

Join me on this journey as we uncover the cutting-edge developments, challenges, and promises that autonomic computing and embodied AI assistants bring to the forefront of artificial intelligence. Together, we will envision a future where technology seamlessly integrates into our lives, making our interactions with machines not just smarter but more human-like than ever before.

VOA using LLMs with Retrieval-Augmented Generation



Experiments indicate that we can accommodate multi-modal inputs (speech + pointing).

Embodied AI

- Vision
- Anatomy
- Prototypes
- **Research challenges**
 - Leveraging Large Language Models
 - **New non-verbal modalities**

Research Challenge: New non-verbal modalities



Image source: <https://www.concentrix.com/blog/workplace-diversity-road-equality-long-good-start/>

- Human-human communication is multi-modal
 - ✓ Speech
 - ✓ Pointing
 - Gesture
 - Head orientation
 - Eye contact
 - Facial expression

With Professor Qiang Ji and students Chenyi Kuang and Yufei Zhang of RPI (Rensselaer Polytechnic Institute), we are developing new non-verbal modalities that are particularly relevant for multi-human scenarios.

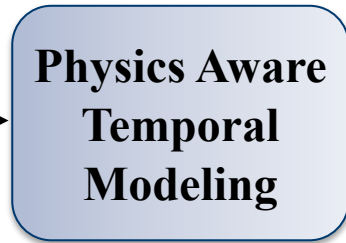
Research challenge

3D reconstruction of body pose from monocular camera

Zhang, Y., Wang, H., Kephart, J. O., & Ji, Q. (2023). Body Knowledge and Uncertainty Modeling for Monocular 3D Human Body Reconstruction. *arXiv preprint arXiv:2308.00799*.



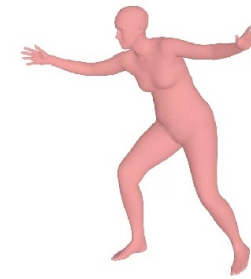
2D Image or Video



$$p(\theta|X; W)$$

$$p(\beta|X; W)$$

Distributions of body
pose and shape



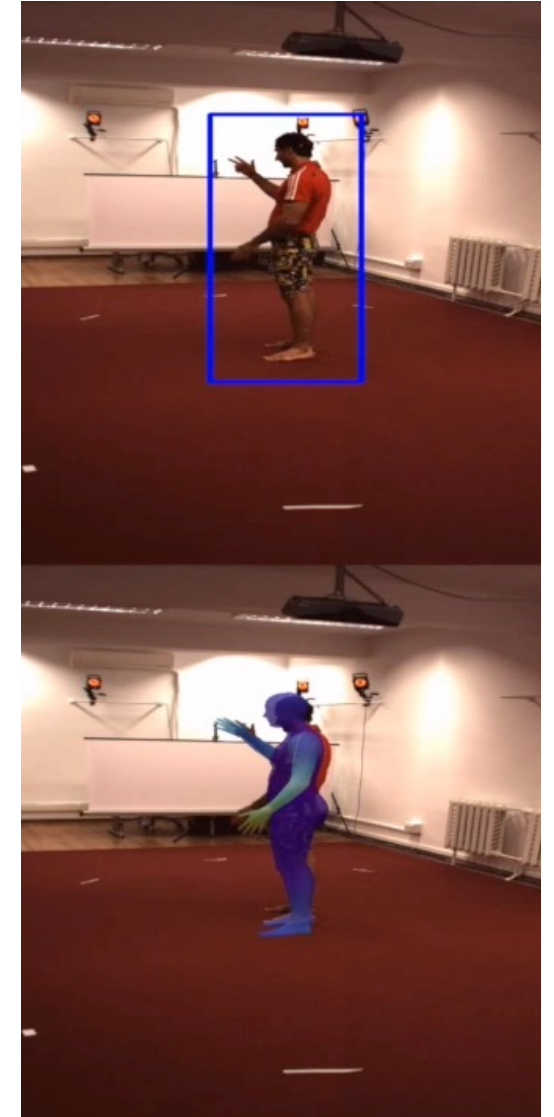
3D body mesh

$$\tau$$

$$\lambda$$

Joint actuation and
Ground reaction force

We can use physics
and bio-mechanical
constraints to infer
3D from 2D

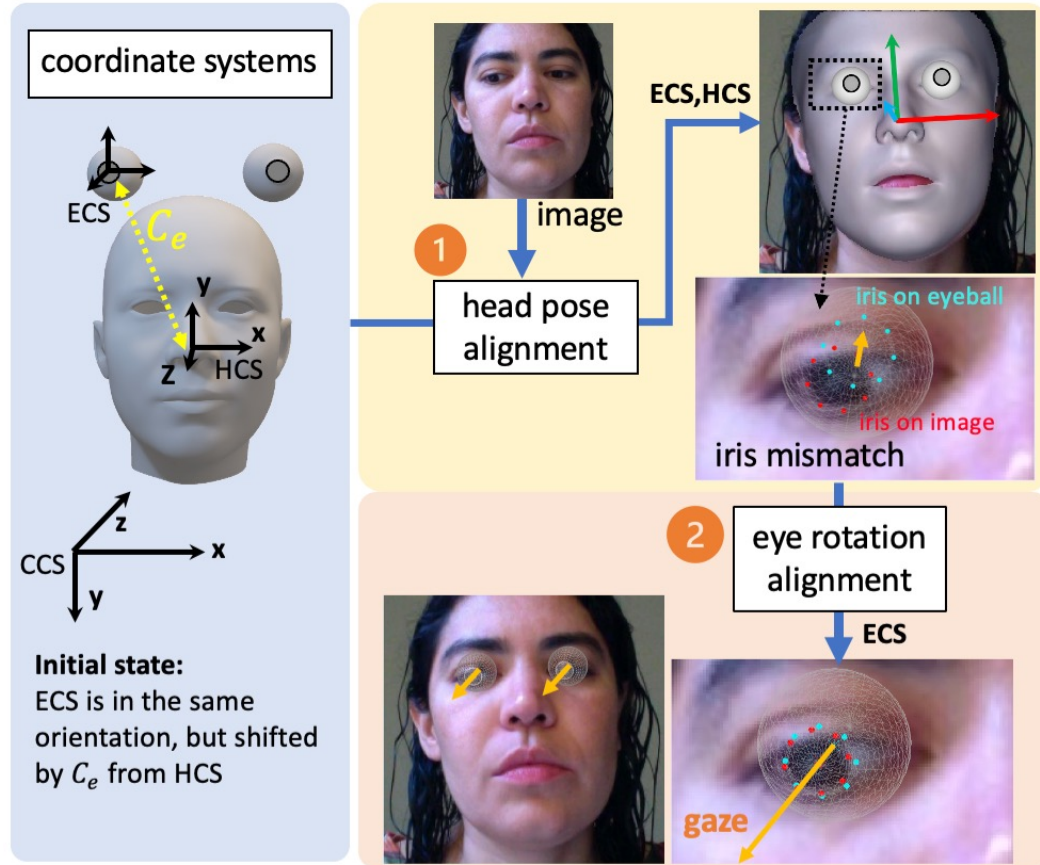


3D reconstruction with
uncertainty visualization

3D Eye Gaze Estimation with eyeball modeling

Monocular camera

Workflow



Real-world demonstration

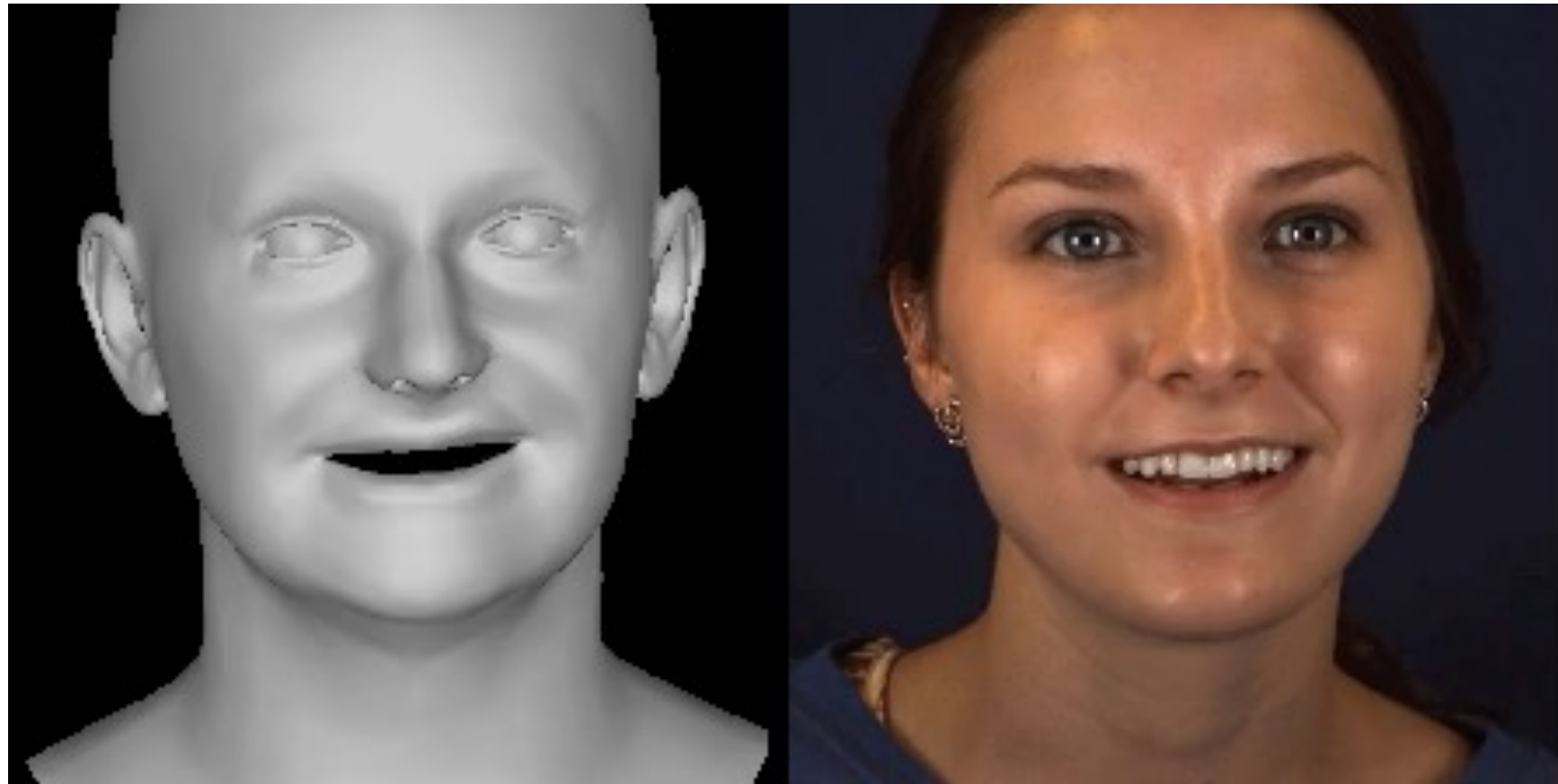


Accuracy: ~5-10 degrees (more difficult with large head rotation)

Dynamic 3D Face Reconstruction

Monocular camera

:



3D Reconstruction
mapped onto neutral face

Original 2D video

Kuang, C., Cui, Z., Kephart, J. O., & Ji, Q. (2022, October). AU-Aware 3D Face Reconstruction through Personalized AU-Specific Blendshape Learning. In *European Conference on Computer Vision* (pp. 1-18). Cham: Springer Nature Switzerland.

We can accurately capture 3D movements of mouth, eyebrows and eyelids.

This can be leveraged to understand facial expressions and the underlying muscle contractions that produce them.

Outline

- Autonomic Computing and Utility Functions
- A fatal flaw
- Embodied AI
- Embodied AI for Autonomic Computing
- Autonomic Computing for Embodied AI
- Conclusions

Embodied AI for System Administration (prototype)

The screenshot displays the M2A2 Dashboard with a chatbot interface on the left and two data tables on the right. The chatbot, named 'celia', is interacting with a user named 'Jeffre...'. The chat history shows the user asking for control scenarios, the chatbot providing 7 results sorted by energy savings, the user asking for the best option, the chatbot providing a decision assistant, the user asking for the best option again, and the chatbot providing another decision assistant.

The first table, titled 'control_scenario-query-table:2023-09-24T03:49:33.936Z', shows 7 matching records for the Control_scenario Table. The columns are Scenario ID, Energy Savings Mode, Maint. Frequency (per day), Servers (Gold), Servers (Silver), % Compliance (Gold), and % Compliance (Silver). The records are as follows:

Scenario ID	Energy Savings Mode	Maint. Frequency (per day)	Servers (Gold)	Servers (Silver)	% Compliance (Gold)	% Compliance (Silver)
0 S21.11.2.1	2	1	21	11	78.1468	86.6220
1 S22.10.2.1	2	1	22	10	83.8135	74.9794
2 S21.11.2.0.5	2	0.5	21			
3 S22.10.2.0.5	2	0.5	22			
4 Current	3	1	21			
5 S22.10.3.0.5	3	0.5	22			
6 S22.10.4.0.5	4	0.5	22			

The second table, titled 'smartswaps:2023-09-24T03:54:39.669Z', is a Smart Swap Decision Table. It allows users to choose between 'Best Choice' and 'Original Value' views. The columns are ScenarioID, EnergySavings, MaintenanceFrequency, ComplianceGold, and ComplianceSilver. The records are as follows:

ScenarioID	EnergySavings	MaintenanceFrequency	ComplianceGold	ComplianceSilver
Current	3	1	73.20651364653867	84.99998561223259
S21.11.2.1	2	1	78.14684150691834	86.62204819586691
S22.10.2.1	2	1	83.81352356949124	74.97936615393752
S21.11.2.0.5	2	0.5	88.49442900136447	98.7873219285184
S22.10.2.0.5	2	0.5	94.85787114841659	84.95046499221766
S22.10.4.0.5	4	0.5	89.67025743923548	74.99974632252082
S22.10.3.0.5	3	0.5	92.94711508951089	74.99999993666407

The Smart Swap Decision Table also includes a 'Recommend' checkbox and a decision prompt: 'Current is superior to S22.10.2.1 in most aspects. Remove S22.10.2.1?'. The prompt includes buttons for 'YES, REMOVE S22.10.2.1.', 'NO, REMOVE CURRENT INSTEAD.', 'UNSKIP', and 'SKIP'.

- What-if analysis
- Simulation
- Decision assistance

[video]

Imagine what additional existing capabilities could be integrated Based on existing technology

The screenshot displays a dashboard with a chatbot interface on the left and data tables on the right. The chatbot is titled 'm2a2-chatbot' and shows a conversation with a user named 'Jeffrey'. The user asks for help choosing the best option, and the chatbot responds with a decision assistant that helps choose among options based on energy savings. The data tables include 'Control Scenario Table' and 'Smart Swap Decision Table'.

Scenario ID	Energy Savings Range	Maint. Frequency (per day)	Server (P04)	Server (S04)	% Compliance (P04)	% Compliance (S04)
S2111.2.1	2	1	21	11	78.1468	86.6220
S22.10.2.1	2	1	21	22		
S2111.2.0.5	2	0.5	21			
S22.10.2.0.5	2	0.5	22			
Current	3	1	21			
S22.10.3.0.5	3	0.5	22			
S22.10.4.0.5	4	0.5	22			

ScenarioID	EnergySavings	MaintenanceFrequency	ComplianceP04	ComplianceS04
Current	3	1	73.20611564815367	84.9998541223159
S21.11.2.1	2	1	78.14684100481834	86.62204819586491
S22.10.2.1	2	1	83.8132256449124	74.9795641093752
S21.11.2.0.5	2	0.5	88.4944290733447	98.7973214283184
S21.10.2.0.5	2	0.5	84.82587114841555	84.6054492011556
S22.10.4.0.5	4	0.5	89.47025742920548	74.9797483222082
S22.10.3.0.5	3	0.5	93.2477122881228	74.979795641093752

- Problem diagnosis
 - Leverage knowledge of system logical and physical infrastructure
 - Retrieve relevant instructions or manuals
 - Knowledge graph is key to interpreting human intent and executing appropriate responses
- What-if analyses and simulations
- Workload forecasts
 - Proactive warnings
 - Resource allocation and acquisition suggestions
- Risk-based decision making
 - Elicit user risk preference
 - Coupled with modeling and simulation

Smart Swaps

- Domain-agnostic
- Guides decision-makers towards solutions that best match preferences
- Breaks a complex decision into a several smaller decisions
 - Absolute dominance
 - Probabilistic dominance
 - Equal attributes
 - Even swaps
- Uses Bayesian principles to infer utility function from user choices

Smart Swap Decision Table

Choose View : Best Choice Original Value

ScenarioID	EnergySavings ↑	MaintenanceFrequency ↑	ComplianceGold ↑	ComplianceSilver ↑
Current	3	1	73.20651364653867	84.99998561223259
S21.11.2.1	2	1	78.14684150691834	86.62204819586691
S22.10.2.1	2	1	83.81352356949124	74.97936615393752
S21.11.2.0.5	2	0.5	88.49442900136447	98.7873219285184
S22.10.2.0.5	2	0.5	94.85787114841659	84.95046499221766
S22.10.4.0.5	4	0.5	89.67025743923548	74.99974632252082
S22.10.3.0.5	3	0.5	92.94711508951089	74.99999993666407

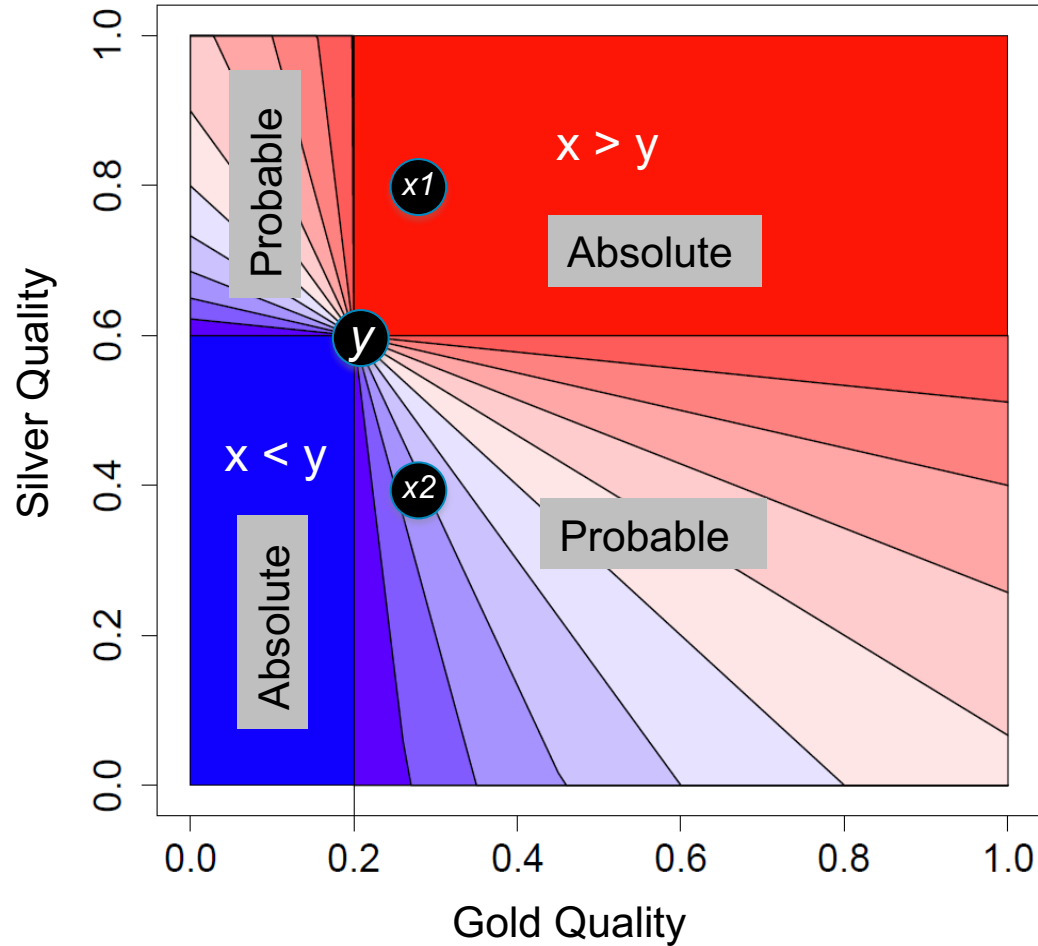
Recommend

ScenarioID	EnergySavings	MaintenanceFrequency	ComplianceGold	ComplianceSilver
Current	3	1	73.20651364653867	84.99998561223259
S22.10.2.1	2	1	83.81352356949124	74.97936615393752

Current is superior to **S22.10.2.1** in most aspects. Remove **S22.10.2.1**?

SmartSwaps: Absolute and Probable Dominance

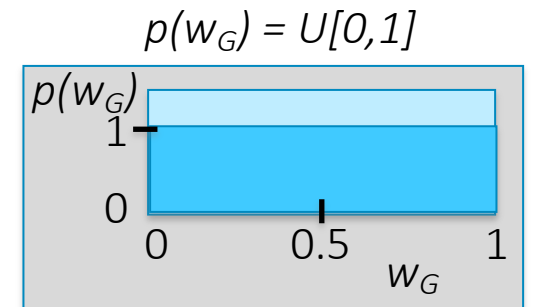
Absolute Dominance



	Gold	Silver
y	0.2	0.6
x_1	0.3	0.8

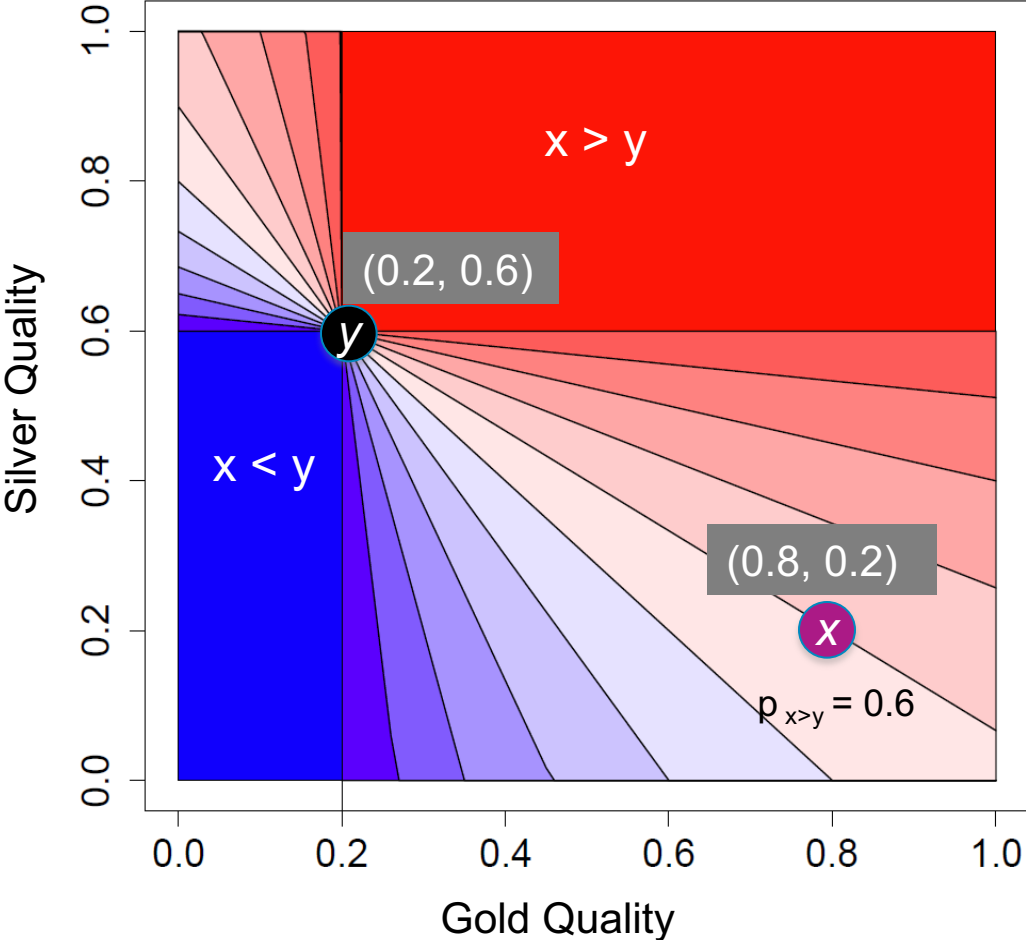
Probable Dominance

	Gold	Silver
y	0.2	0.6
x_2	0.3	0.4

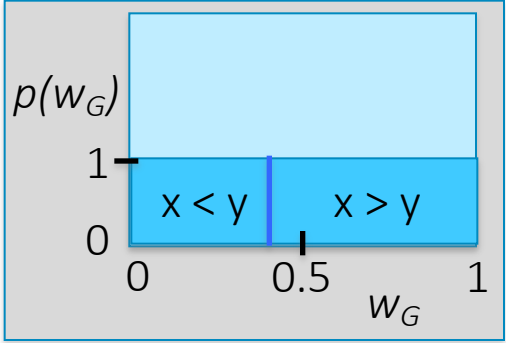


Learning from Probable Dominance

Beliefs about weights: $w_G \sim \text{Uniform}(0,1)$

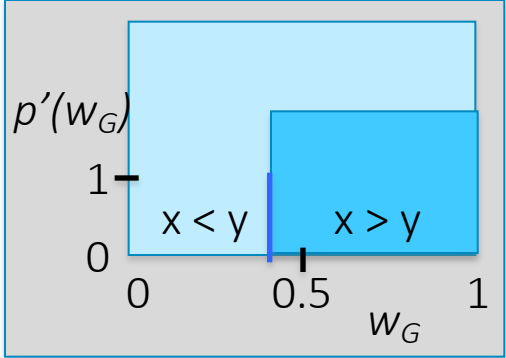


Original beliefs:

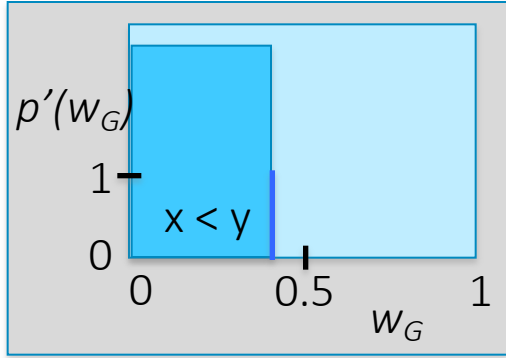


- Suppose we ask user: Is $x > y$?

• If Yes, then



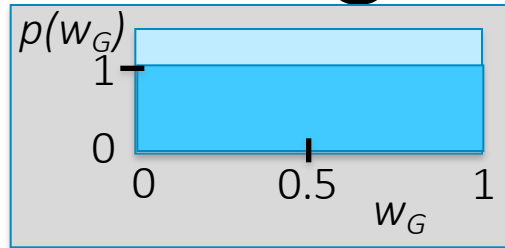
• If No, then



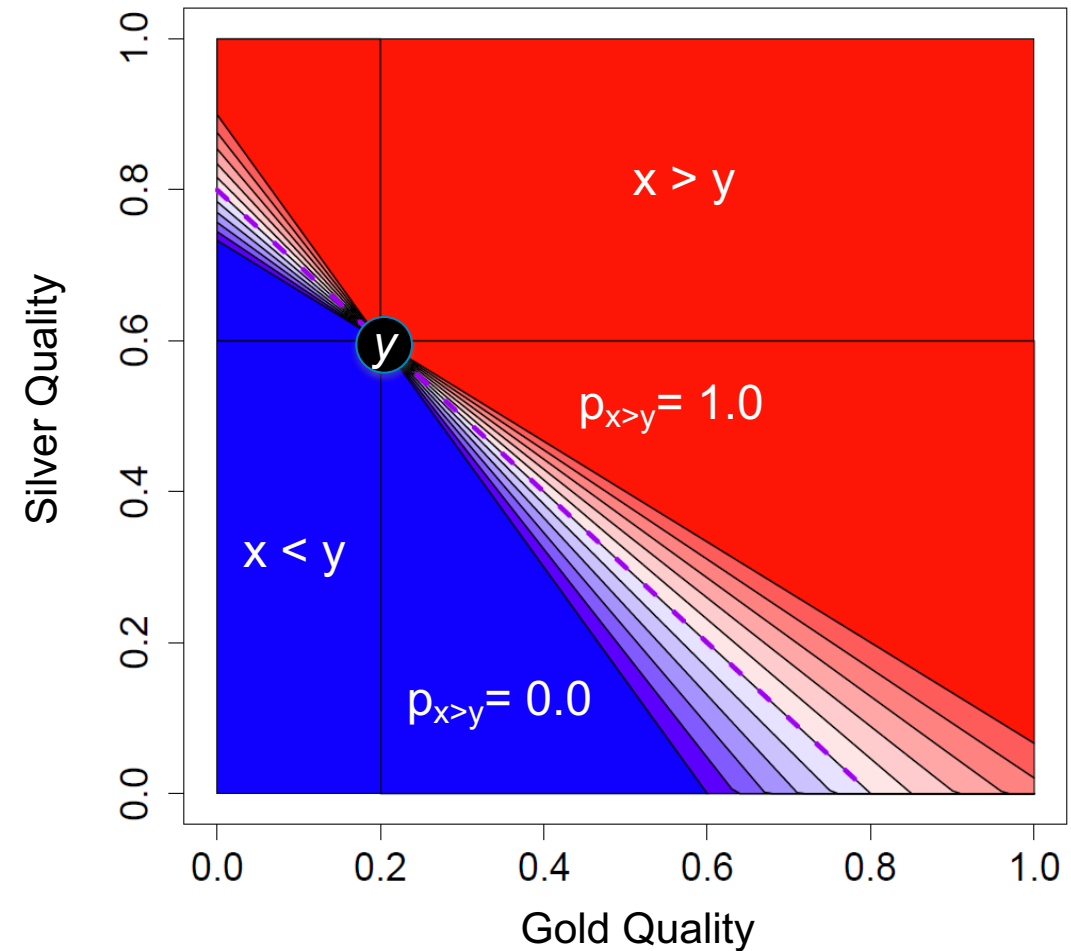
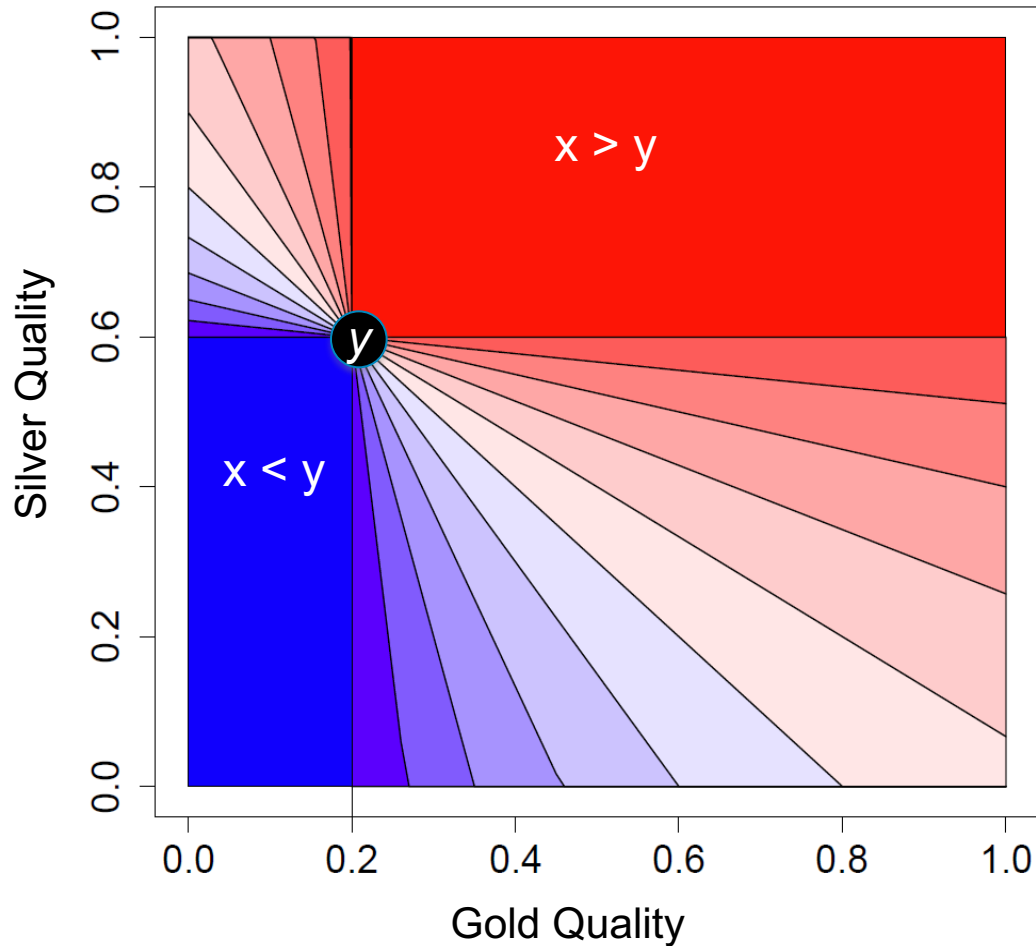
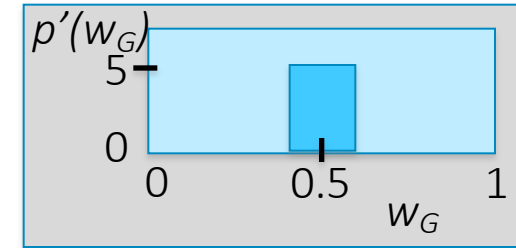
Effect of Learning

SmartSwaps learns the user's utility function!

$$p(w_G) = U[0,1]$$



$$p'(w_G) = U[0.4, 0.6]$$



Outline

- Autonomic Computing and Utility Functions
- A fatal flaw
- Embodied AI
- Embodied AI for Autonomic Computing
- **Autonomic Computing for Embodied AI**
- Conclusions

Autonomic Computing problems for Embodied AI

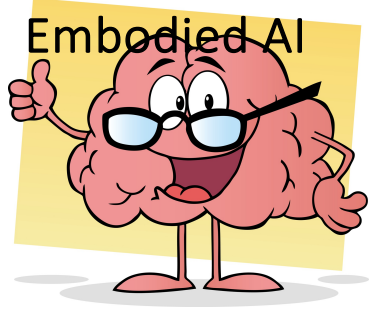
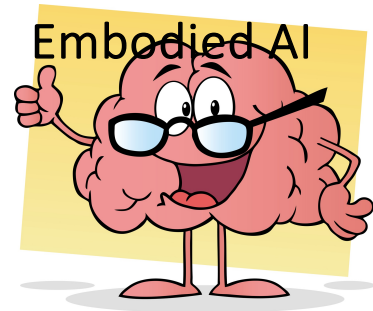
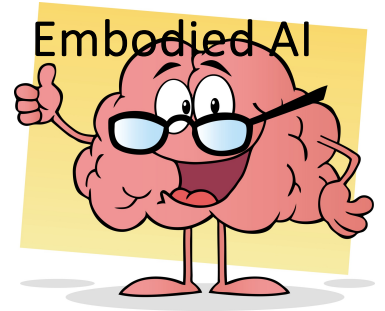
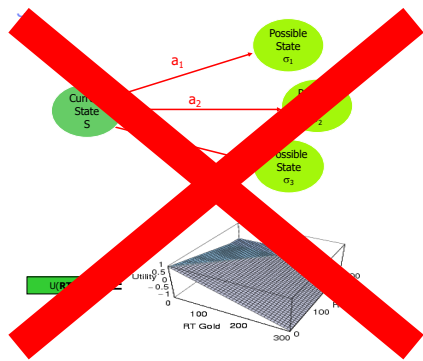
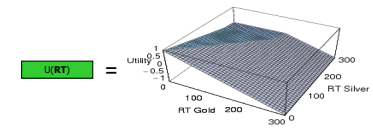
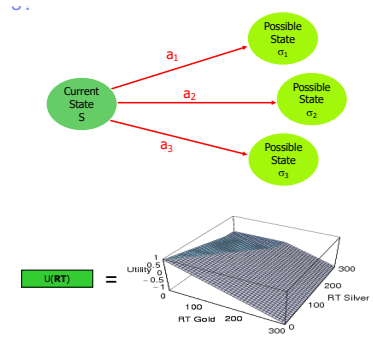
- Embodied AI systems have stringent real-time requirements
 - **High frame rate.** Infer head orientation, gestures, faces, people locations, speech at ~60 frames/sec
 - **Low latency.** To fuse multiple modalities properly, we need accurate timestamping and low latency (< 15-20 msec)
 - **Fluctuating demand.** Demand can be highly sporadic and situation-dependent
 - Proportional to number of humans in the space
 - Dependent on how frequently humans are changing position and orientation, more expensive face recognition can be interspersed with lower-cost tracking technologies.
 - LLMs consume vast GPU resources for training and inferencing
- Embodied AI systems require both edge and cloud computing, and the boundary may shift as demand or technology changes
- We need autonomic computing to embrace these challenges

Outline

- Autonomic Computing and Utility Functions
- A fatal flaw
- Embodied AI
- Embodied AI for Autonomic Computing
- Autonomic Computing for Embodied AI
- **Conclusions**

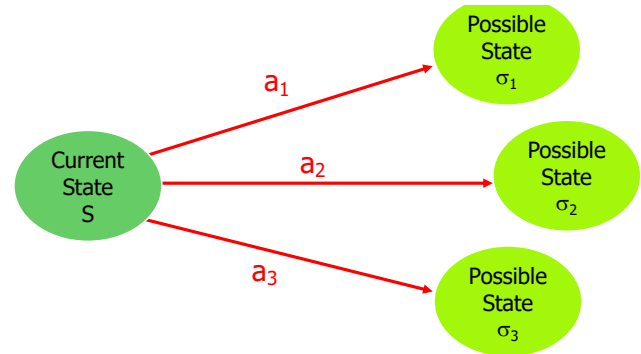
Recap

- AC and Utility Functions
- A fatal flaw
- Embodied Artificial Intelligence
- Embodied AI for AC
- AC for Embodied AI



Conclusions

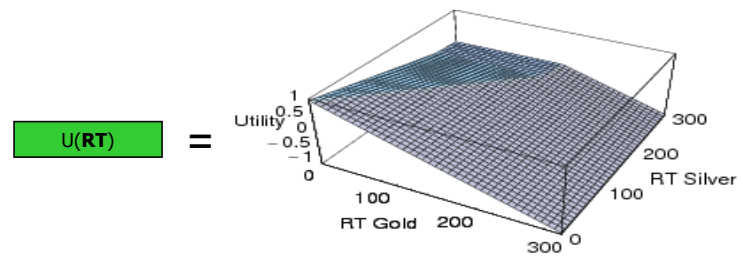
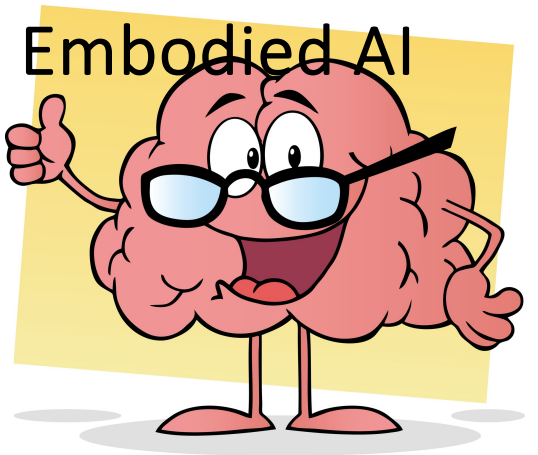
Desired goal state =



The Vision of Autonomic Computing



+



A partnership between autonomic computing and embodied AI

Backup

SmartSwaps evaluation

- Experimental setup

- M attributes
- N alternative choices
- Randomly generate 100 decision tables
- Randomly draw user's true utility weights from (M-1)-dimensional unit simplex
- Record average number of queries of each type over the set of 100

- Results

- Absolute dominance becomes less helpful/important as M grows
- Can reach decision for reasonable-sized problems with ~10 queries

