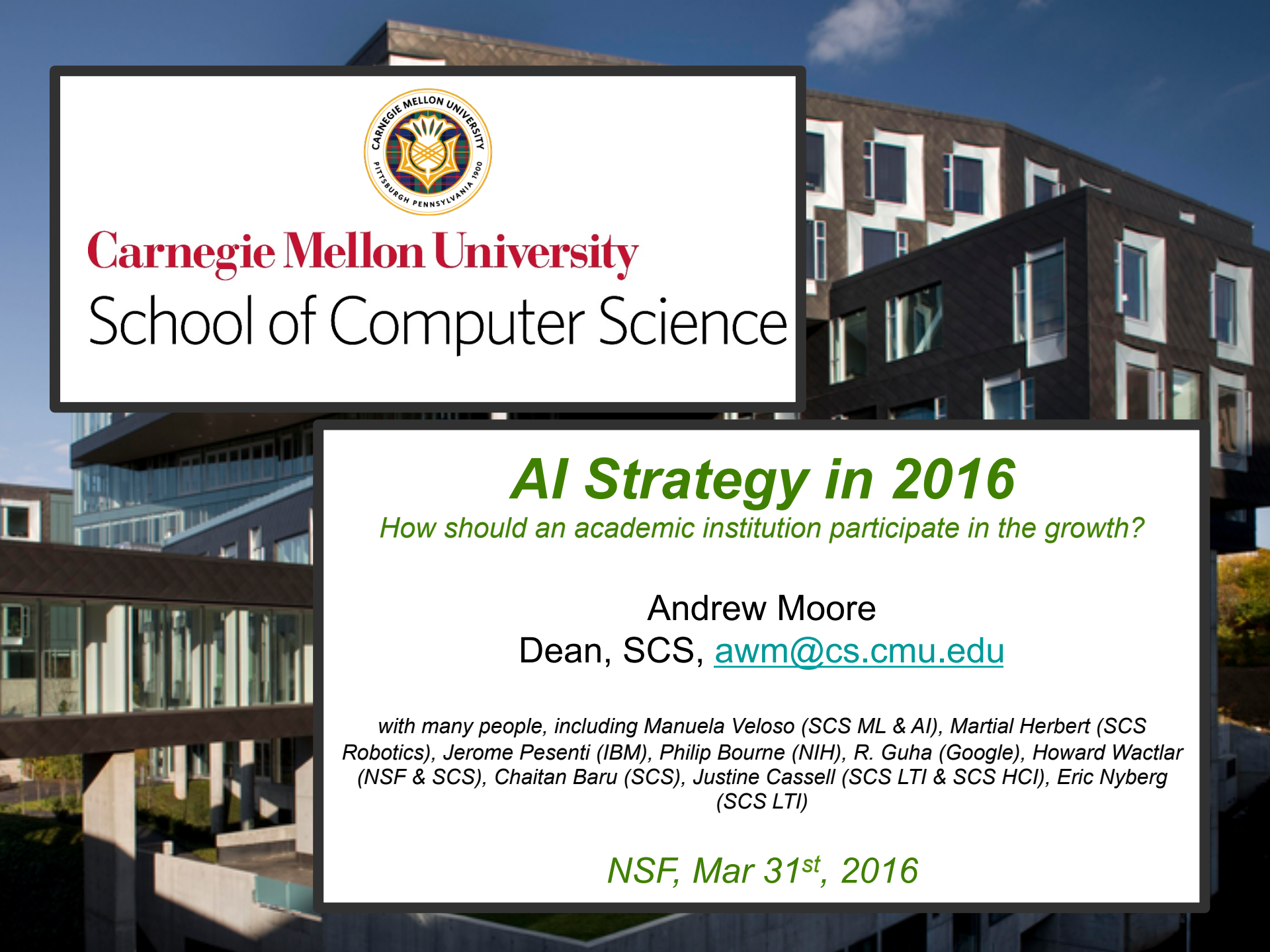


Carnegie Mellon University
School of Computer Science

AI Strategy in 2016

How should an academic institution participate in the growth?

Andrew Moore
Dean, SCS, awm@cs.cmu.edu

with many people, including Manuela Veloso (SCS ML & AI), Martial Herbert (SCS Robotics), Jerome Pesenti (IBM), Philip Bourne (NIH), R. Guha (Google), Howard Wactlar (NSF & SCS), Chaitan Baru (SCS), Justine Cassell (SCS LTI & SCS HCI), Eric Nyberg (SCS LTI)

NSF, Mar 31st, 2016

What's changed since 2011?

What's changed since 2011?

- Data science stack is maturing
- Speech and NLP are becoming commodity
- Highly profitable search engine and consumer electronics business models are serious adopters.
- In industry: there is serious disruption in the air.
- In society: there is dawning excitement and apprehension

What’s changed since 2011?

- Data science stack is maturing
- Speech, vision and NLP are becoming commodity
- Highly profitable search engine and ... consumer electroni... re serious ac...
- In industry ... the air.
- In society: ... xcitement and appreh... nsion

This talk: What’s next, and what roles should major tech universities play?

What's changed since 2011?

- Data science stack is maturing
- Speech, vision and NLP are becoming commodity
- Highly profitable search engine and consumer electron… re serious ac…
- In industry … the air.
- In society: … and apprehension

This talk: What's next, and what roles should major tech universities play?

..in education
..in society
..in technology

What's changed since 2011?

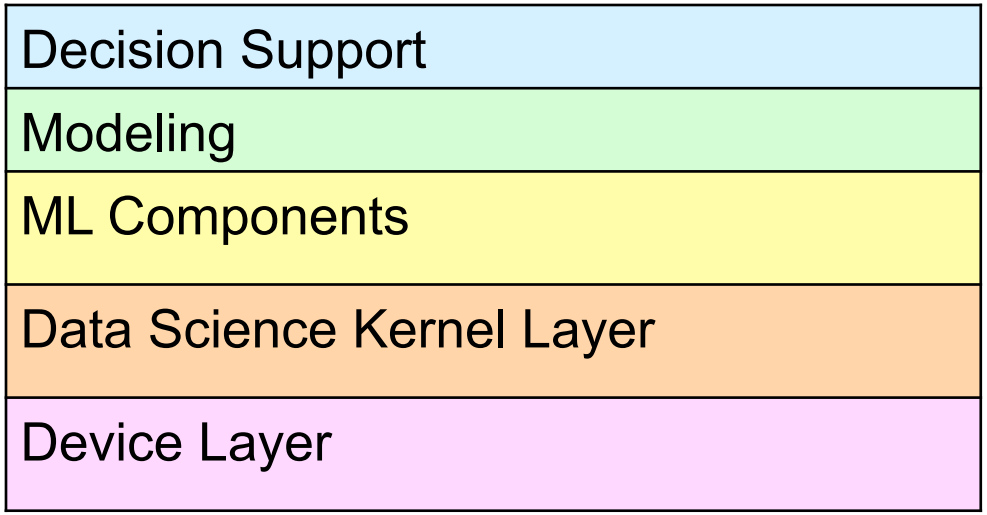
- Data science stack is maturing
- Speech, vision and NLP a… commodity
- Highly profitable search engine and consumer electroni… re serious ac…
- In industry… the air.
- In society: … and apprehension

Let's begin here

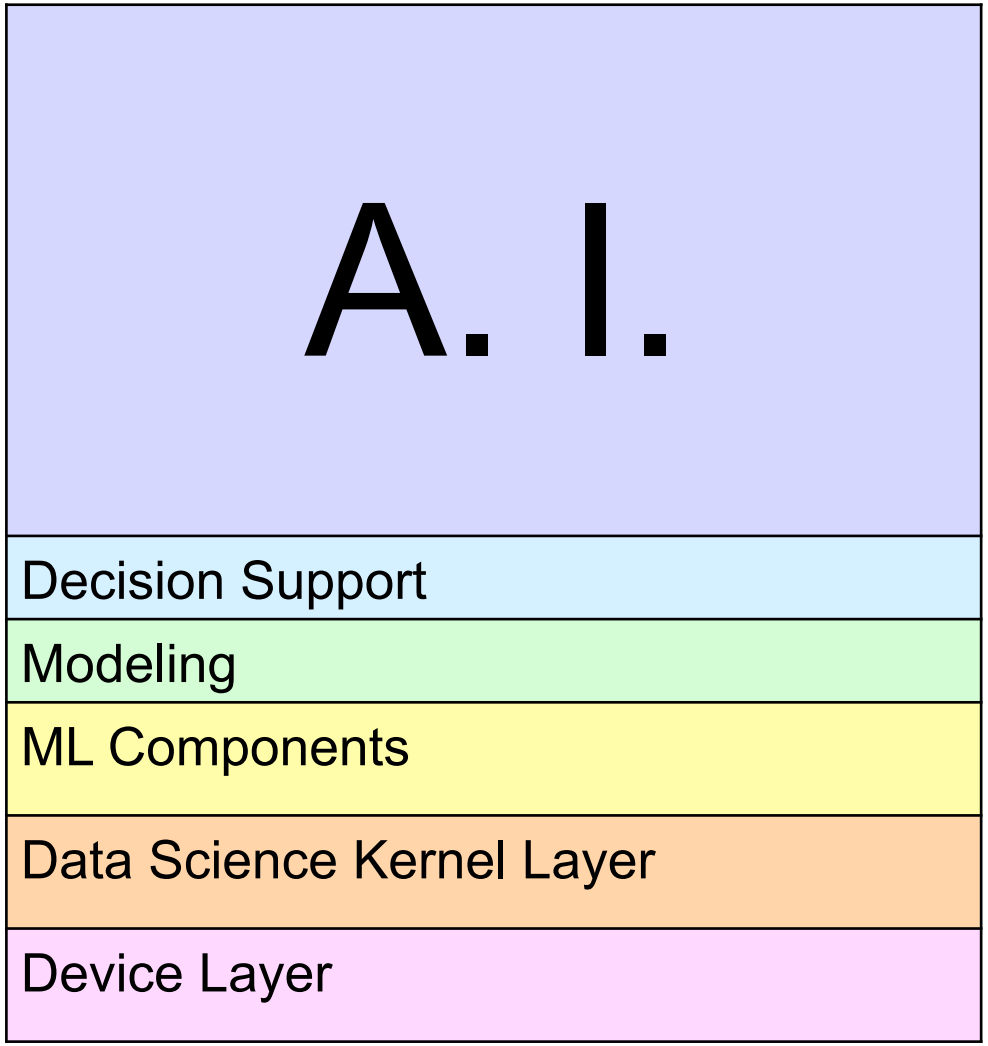
This talk: What's next, and what roles should major tech universities play?

..in education
..in society
..in technology

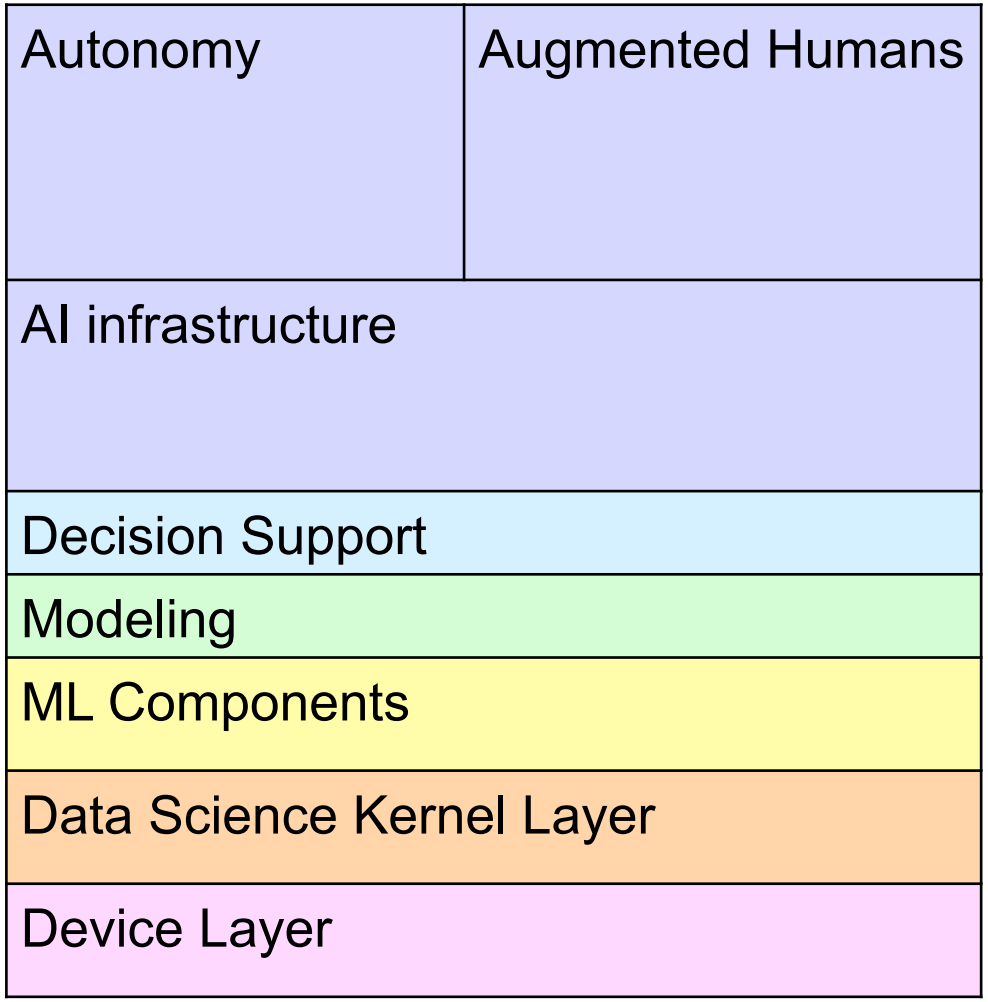
The Data Science Stack



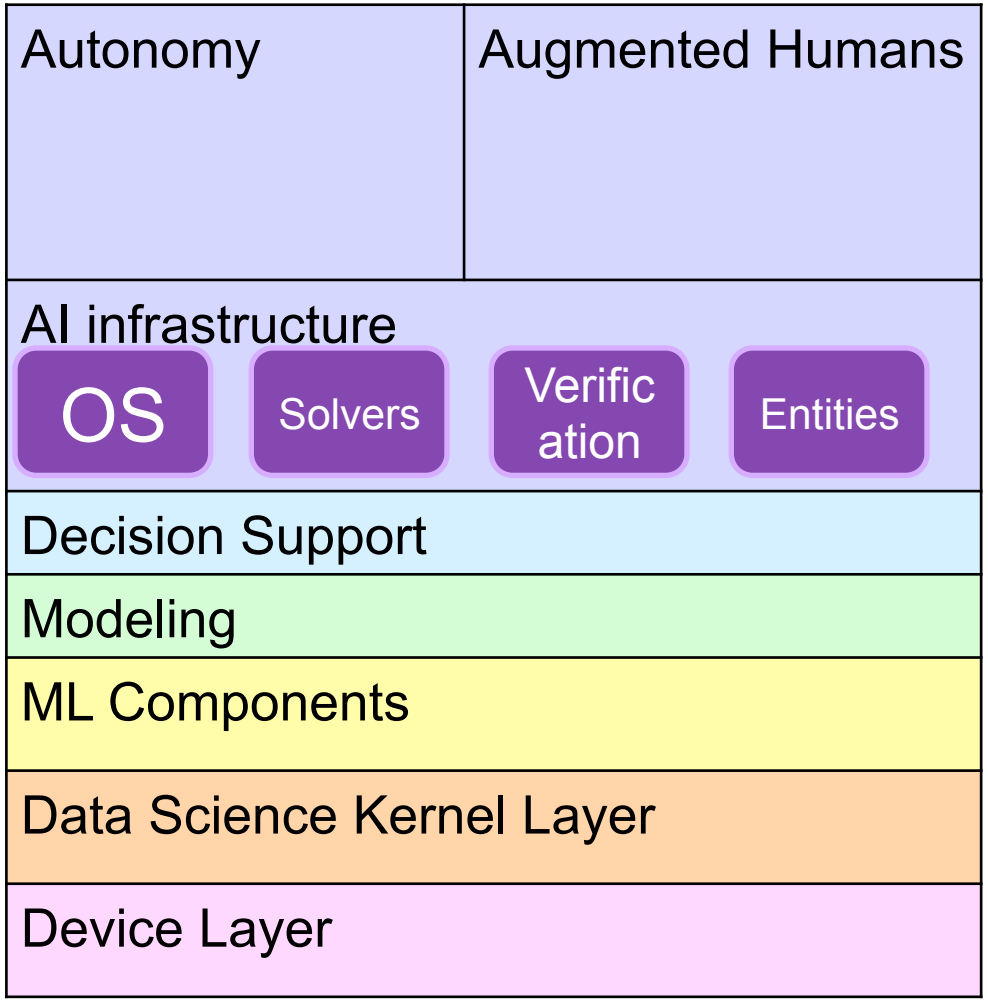
The AI Stack


A. I.
Decision Support
Modeling
ML Components
Data Science Kernel Layer
Device Layer

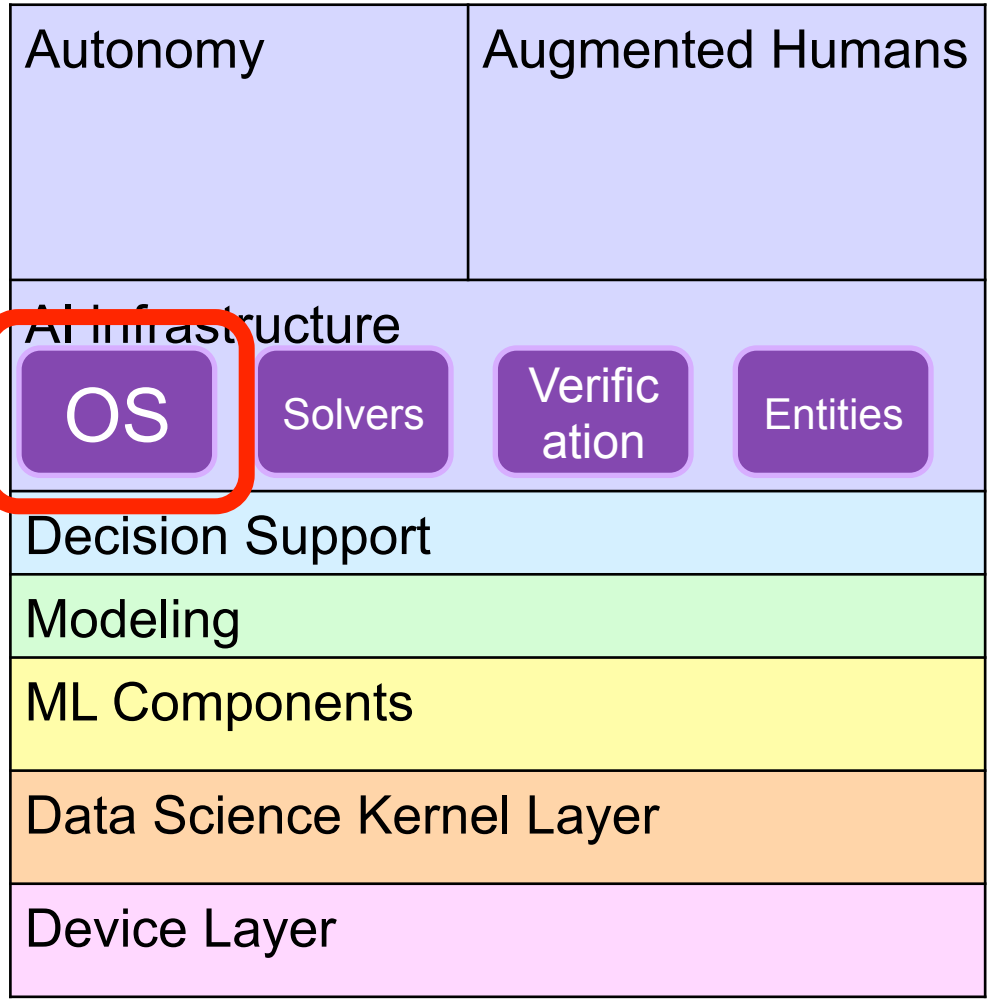
The AI Stack

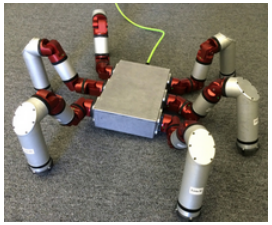

Autonomy
Augmented Humans
AI infrastructure
Decision Support
Modeling
ML Components
Data Science Kernel Layer
Device Layer

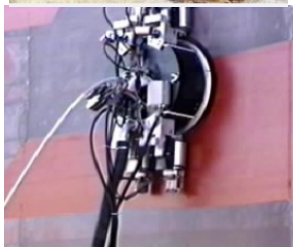
AI infrastructure



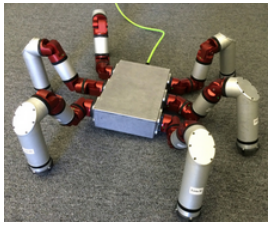
AI infrastructure



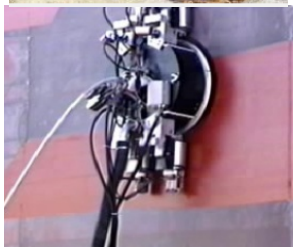




Carnegie Mellon University
School of Computer Science















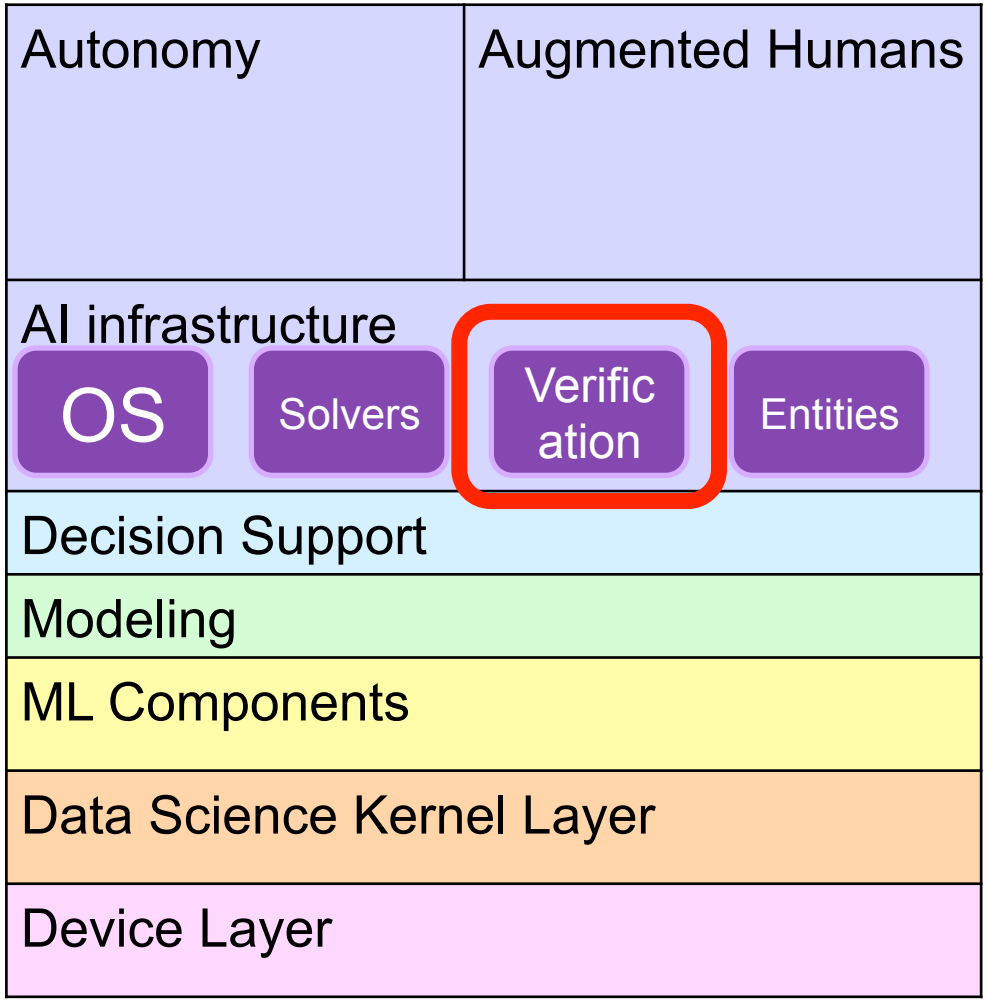




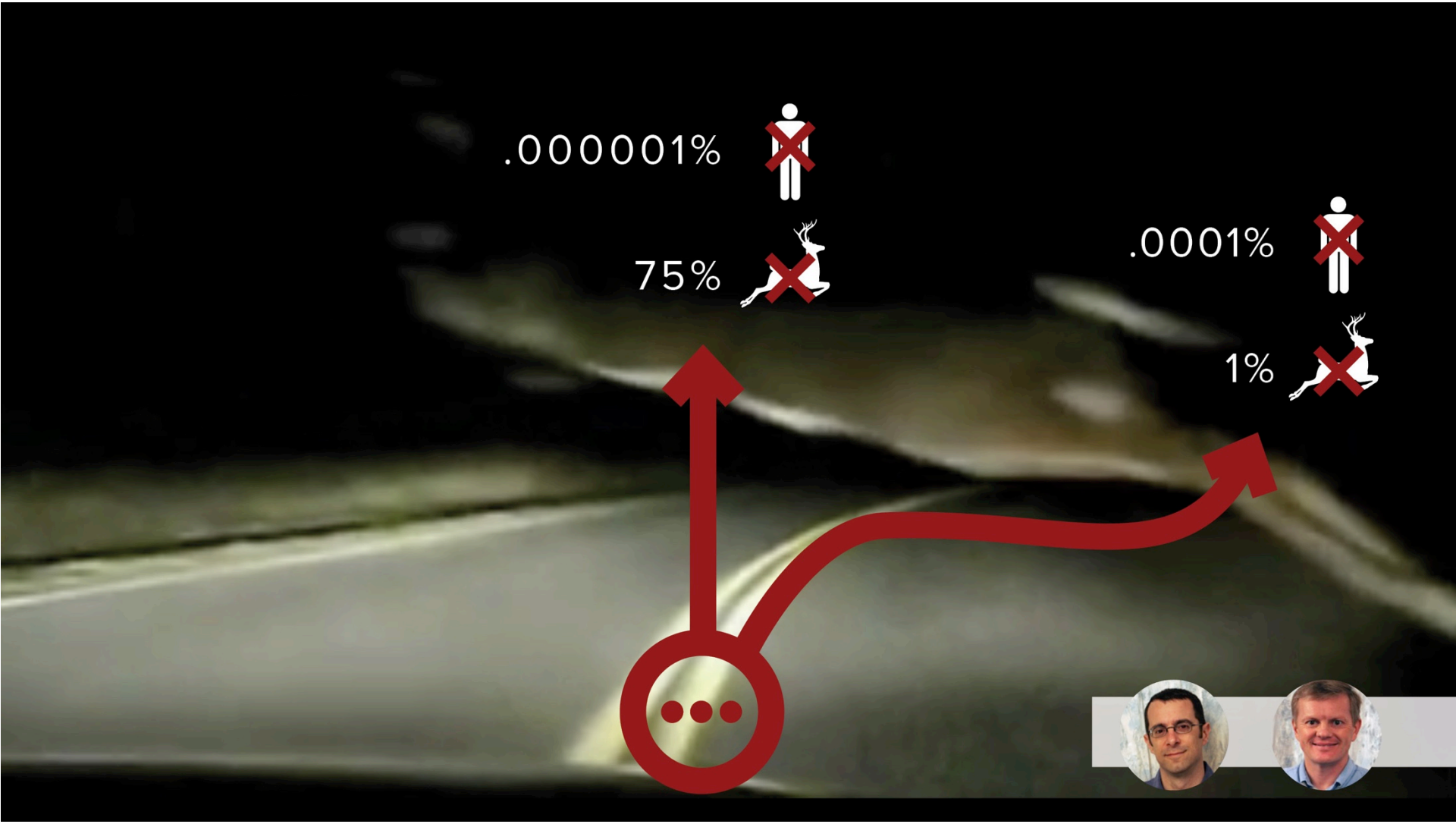
- We build new robot platforms rapidly.
- Startup-company speed, not big engineering speed.
- This is because of know-how, systems and workflows.
- It's time to turn our workflows and software platforms into the world's major robotics operating system

Carnegie Mellon University
School of Computer Science

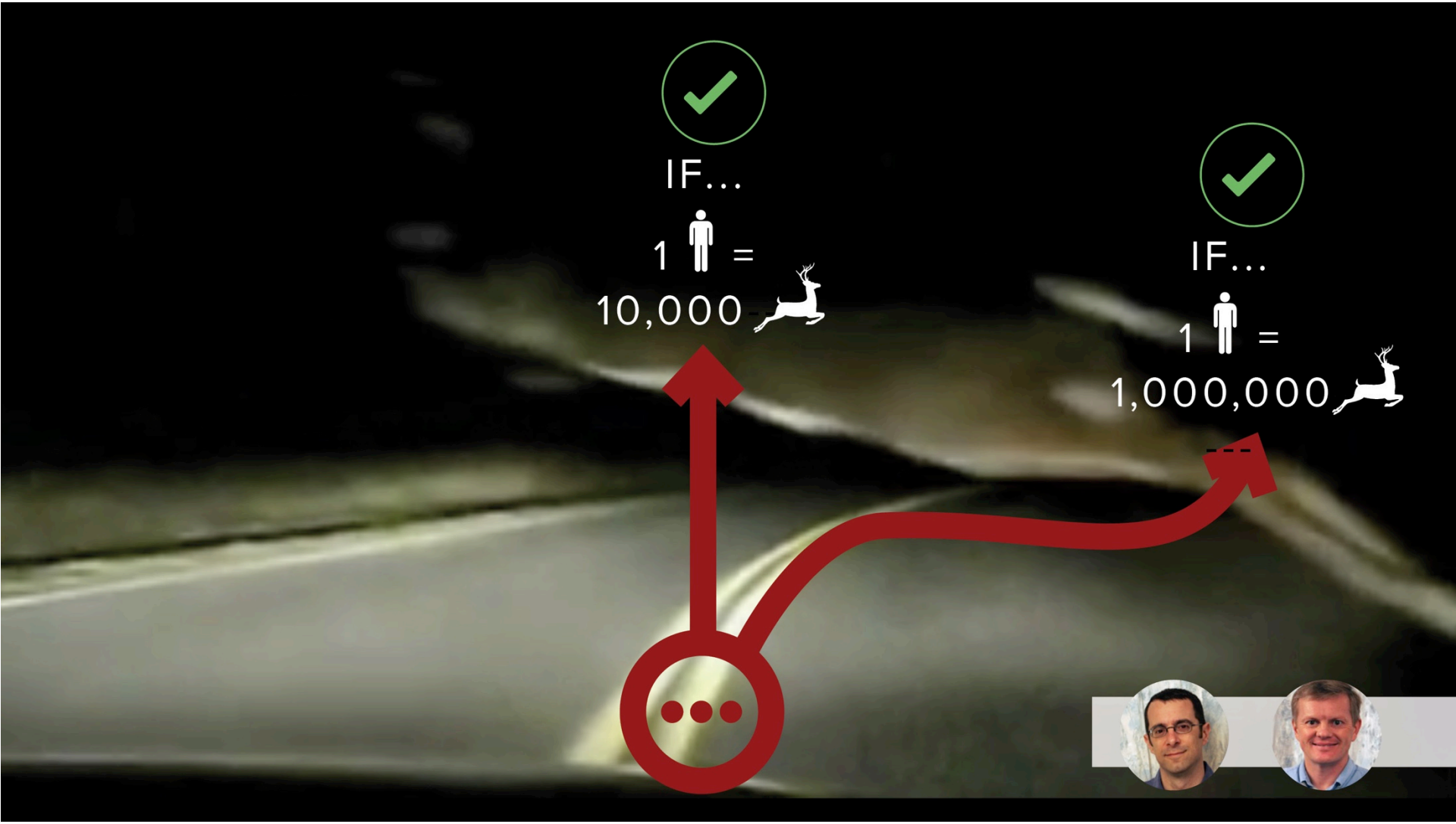
AI infrastructure



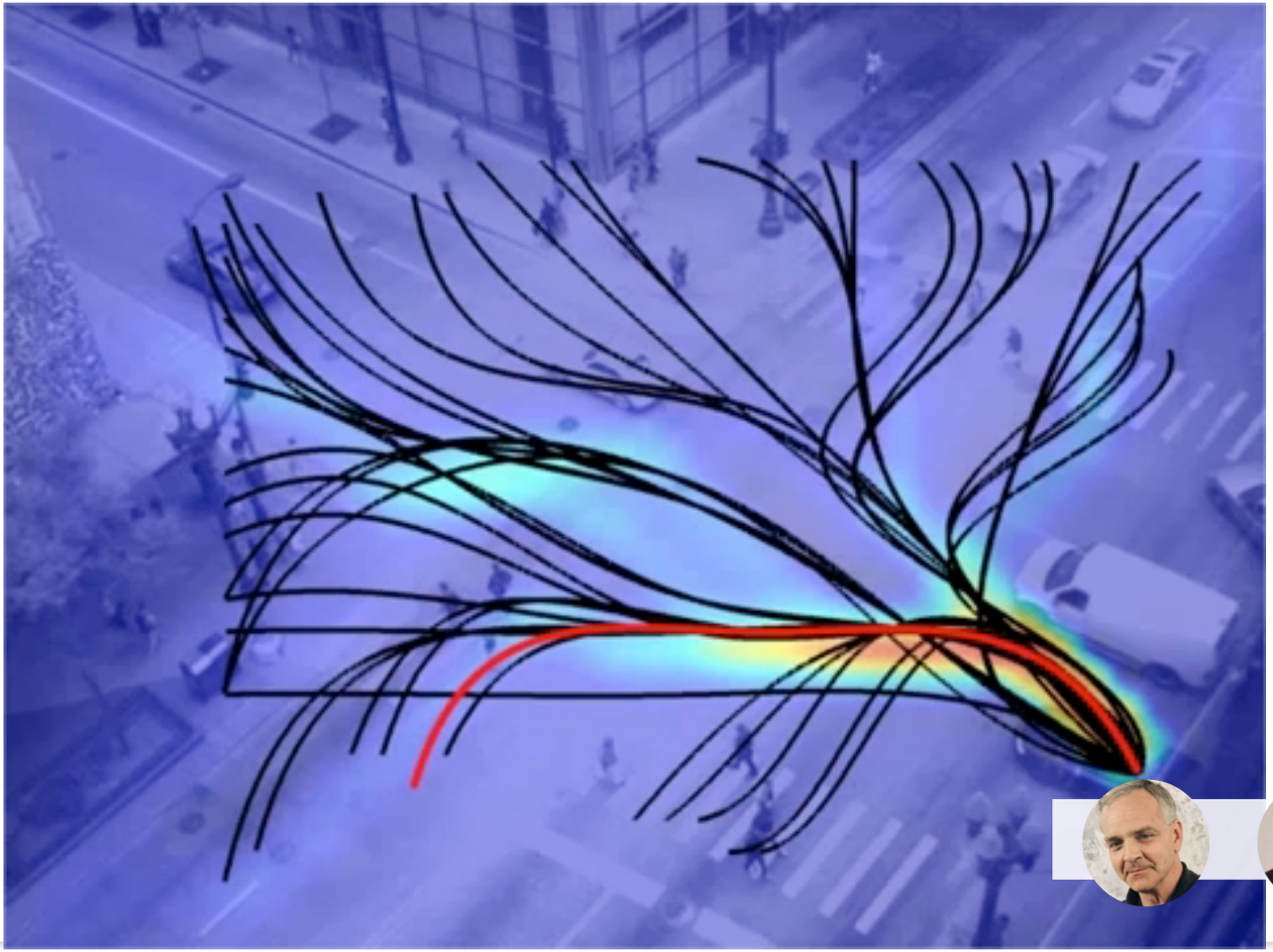
Carnegie Mellon University
School of Computer Science



Carnegie Mellon University
School of Computer Science

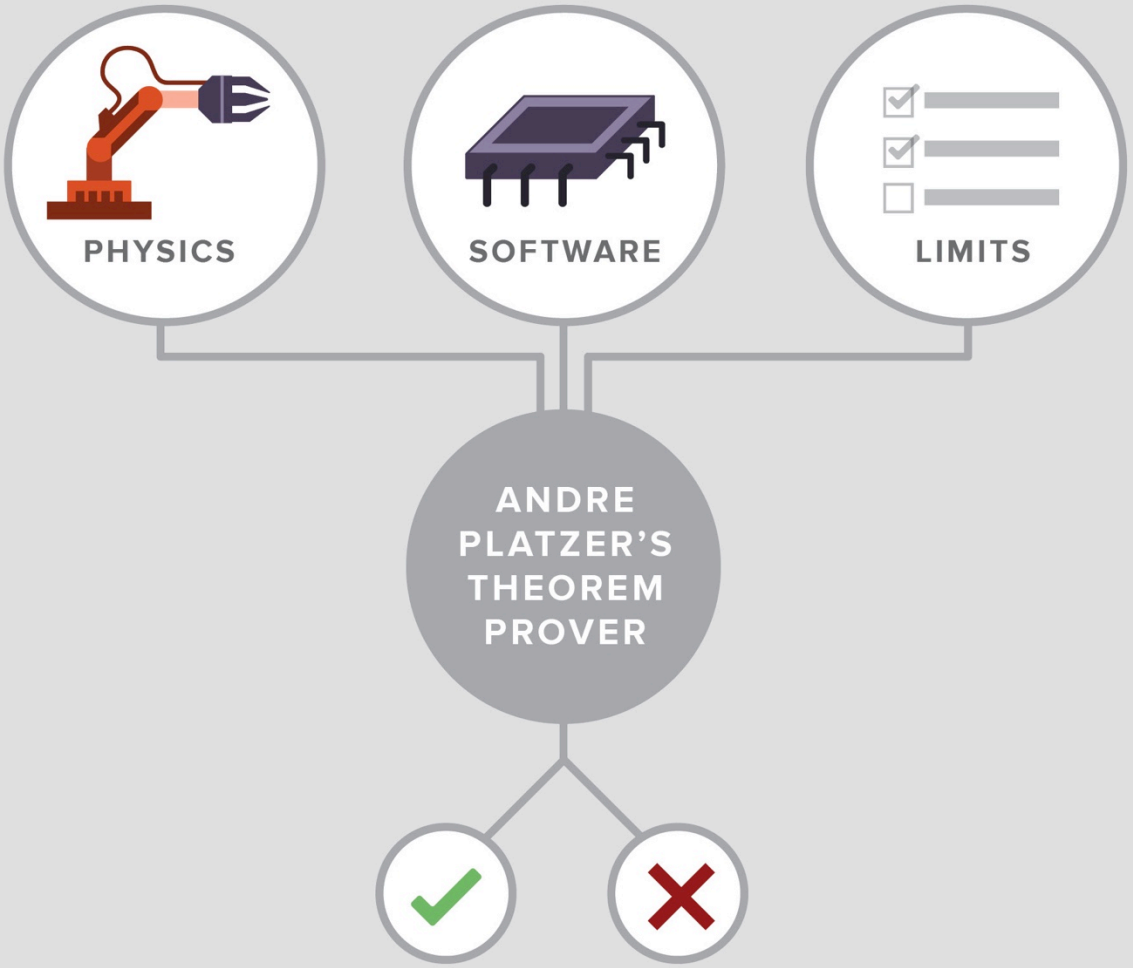


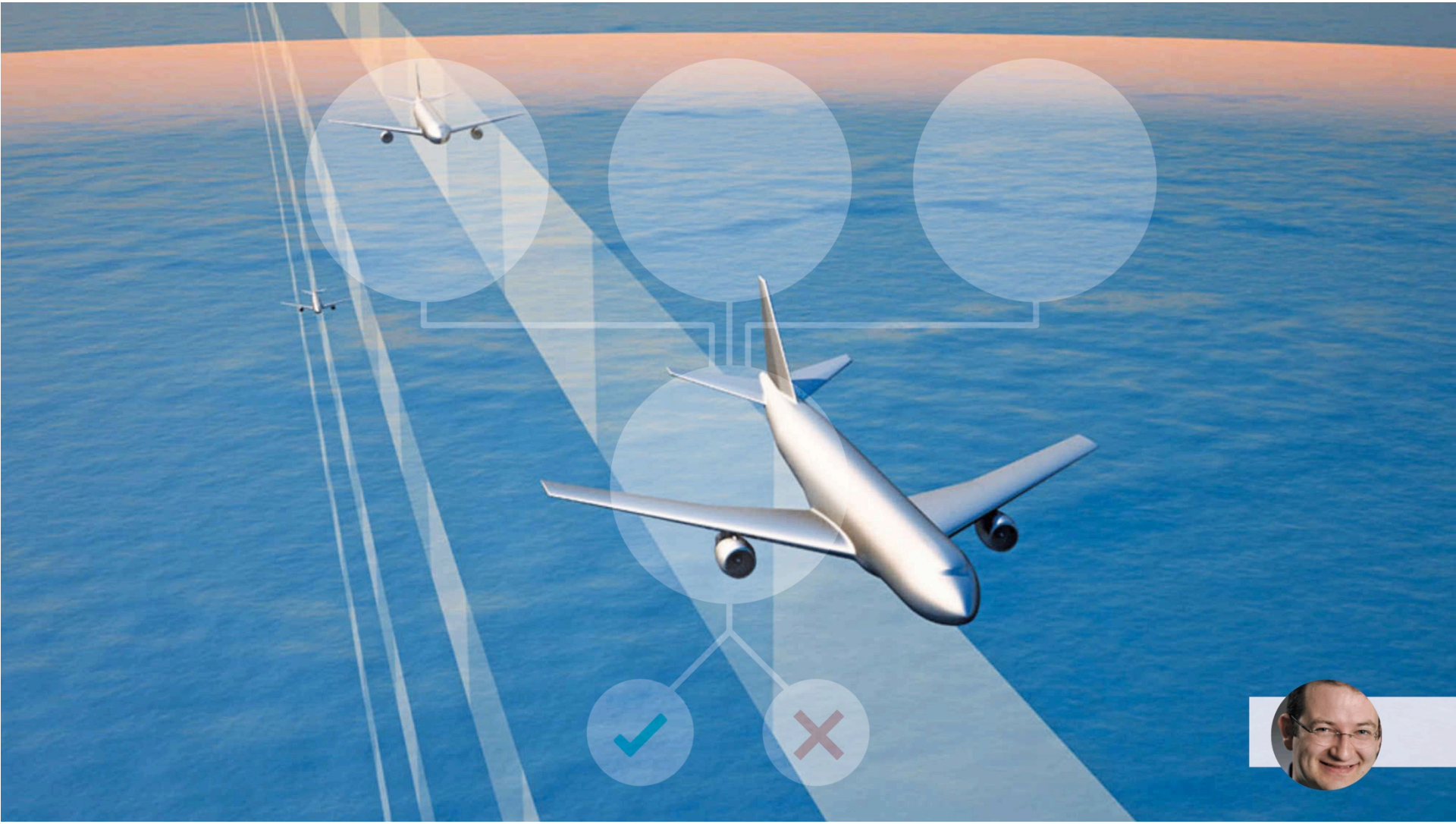
Carnegie Mellon University
School of Computer Science



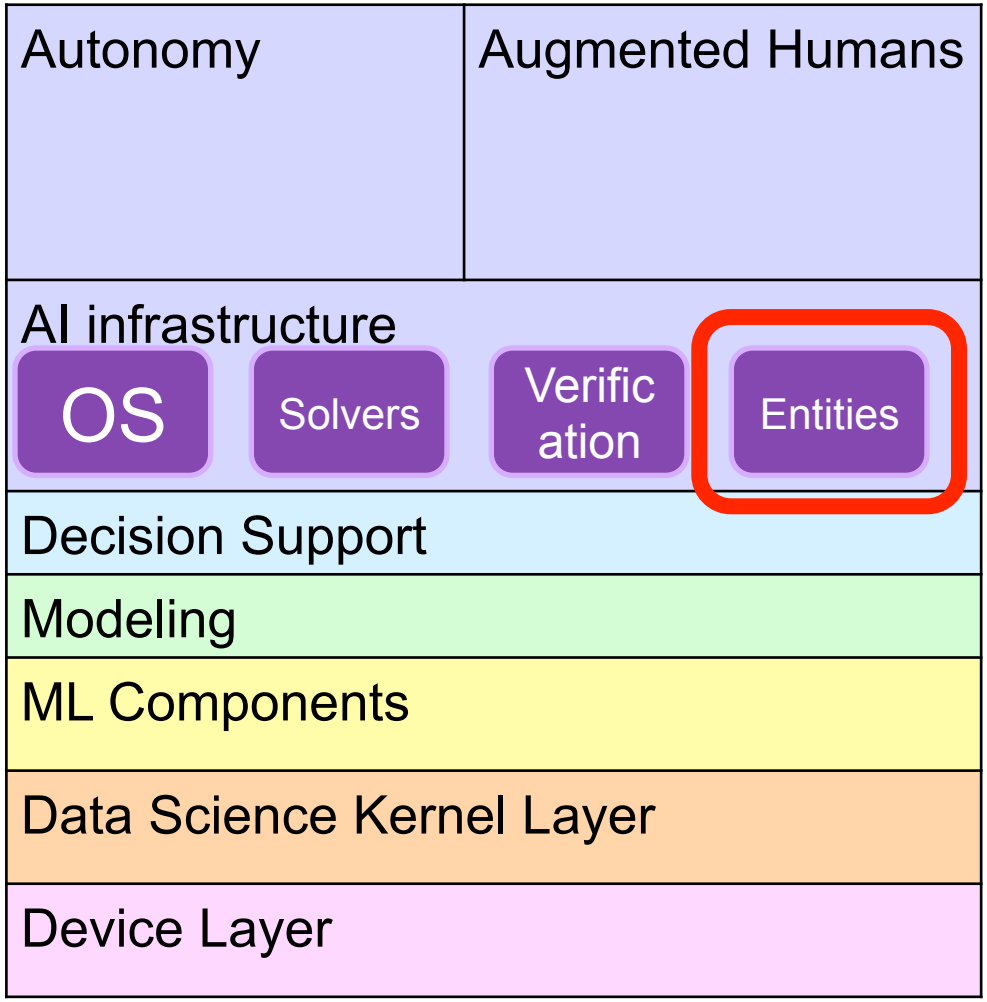
1980	2015
Test 50x	Test 50x
in **16** scenarios	in **1,000,000,000,000,000,000,000,000,000,000,000,000,000,00**


MASTERMIND


PHYSICS
SOFTWARE
LIMITS
ANDRE PLATZER'S THEOREM PROVER



AI infrastructure



- What's an Entity Store?
- Existing Entity Stores
- Architecture
  - Catalog
  - Matching Engine
  - Facts
- Use Cases
- Building a global entity store (internet company view)
- Risks
- Next Steps

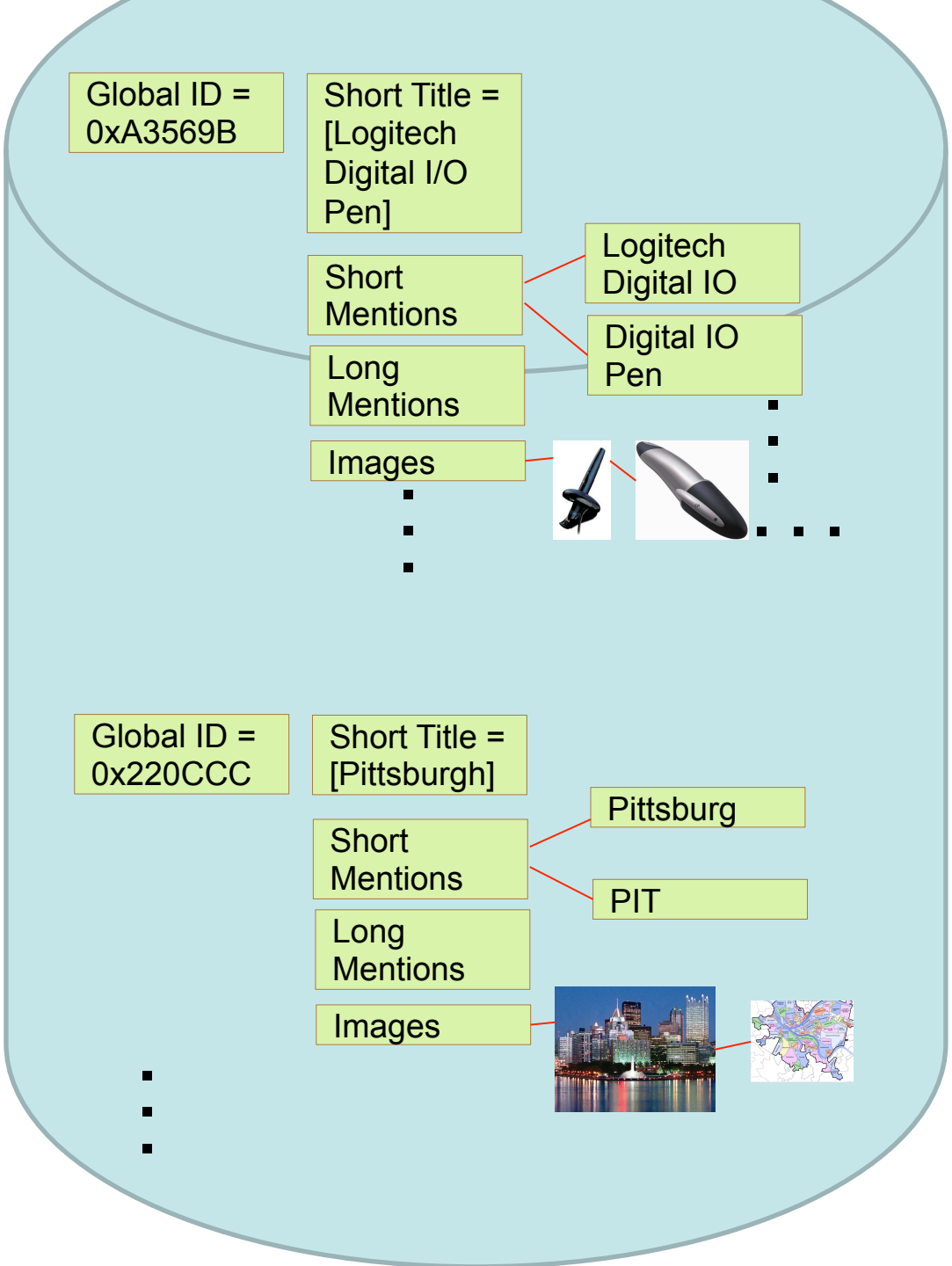
- What's an Entity Store?
- Existing Entity Stores
- Architecture
  - Catalog
  - Matching Engine
  - Facts
- Use Cases
- Building a global entity store (internet company view)
- Risks
- Next Steps

- An entity is a concept from
  - Business
  - Science
  - Administration
  - Entertainment
  - Others
- Examples:
  - The Eiffel Tower
  - Patent #384576
  - Department of Music in College of Fine Arts at CMU
  - The H2N2 Virus
  - The 2011 H2N2 outbreak in Taiwan
  - Logitech digital IO Pen
- A Global Entity store would contain trillions of entities and relations between them

- What's an Entity Store?
- Existing Entity Stores
- Architecture
  - Catalog
  - Matching Engine
  - Facts
- Use Cases
- Building a global entity store (internet company view)
- Risks
- Next Steps

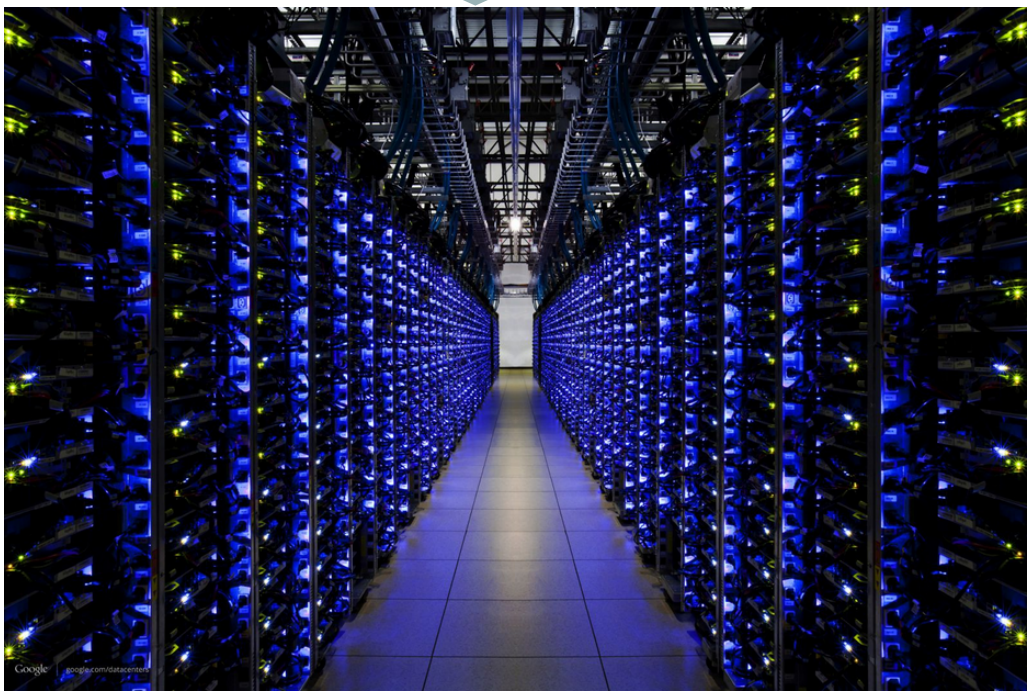
- GIS
- ONR Division 311 KBs
- Amazon
- UMLS Codes
- CYC factbase
- freebase
- tripadvisor etc
- schema.org

- What's an Entity Store?
- Existing Entity Stores
- Architecture
  - Catalog
  - Matching Engine
  - Facts
- Use Cases
- Building a global entity store (internet company view)
- Risks
- Next Steps



- What's an Entity Store?
- Existing Entity Stores
- Architecture
  - Catalog
  - Matching Engine
  - Facts
- Use Cases
- Building a global entity store (internet company view)
- Risks
- Next Steps

"….unlike Logi-Techs' new digital IO stylus, which…"



Global ID = 0xA3569B with probability 0.94
Global ID = 0xEEA001 with probability 0.02
…

- What's an Entity Store?
- Existing Entity Stores
- Architecture
  - Catalog
  - Matching Engine
  - Facts
- Use Cases
- Building a global entity store (internet company view)
- Risks
- Next Steps

"Triples" is one popular approach:

- <Banana ID>.color = <yellow ID>
- (<Barak Obama ID> is_a <Politician ID>)
- (<Dell XPS 13" notebook 2015 ID> has_a <2mm 12 Volt DC composite power socket ID>)

- There is and will continue to be a major intellectual war on the expressiveness of the semantics.
- Winner should be decided by use cases.

- What's an Entity Store?
- Existing Entity Stores
- Architecture
  - Catalog
  - Matching Engine
  - Facts
- Use Cases
- Building a global entity store (internet company view)
- Risks
- Next Steps

Question answering:

Fact Questions:

- *[How old is president Obama?]*
- *[Which Washington-based think tanks have worked on projects involving South American trade?]*
- *[Which building am I in? Where do I go for a taxi?]*

Research Questions:

- *[What are good things to do with kids in Pittsburgh?]*
- *[Which Hodgkins Lymphoma treatments are covered under the Affordable Care Act for my mother?]*
- *[What do the cells in capillary systems of liver tumors unresponsive to <compound> have in common?]*

Fast help for analysts and decision makers

An executive or a scientist or an analyst wishes to unconsciously perform joins with data she is using.

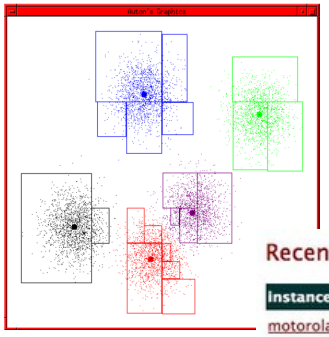
Knowledge-powered robotics

Common sense reasoning; a robot needs to understand, not simply sense, its environment.

Knowledge powered startup and app developer ecosystem

Generally making it easier to write a useful app for domain *X* which needs to know about entities in domain *Y* (e.g. a great supply center navigation app actually needs to know the meanings of the diseases treated by supplies in meds section).

- What's an Entity Store?
- Existing Entity Stores
- Architecture
  - Catalog
  - Matching Engine
  - Facts
- Use Cases
- Building a global entity store (internet company view)
- Risks
- Next Steps



Recently-Learned Facts twitter Refresh

Instance	Iteration	Date Learned	Confidence
motorola a680 is a product	703	27-feb-2013	97.5
guitar service is a musical instrument	703	27-feb-2013	93.8
robert lafleur airport is an airport	703	27-feb-2013	96.9
tomato taco soup crock pot is a vegetable	704	28-feb-2013	90.2
garrett jones is an Australian person	703	27-feb-2013	93.8
tim robbins has wife susan sarandon	708	07-mar-2013	93.8
elephant is an animal that is a kind of animals	704	28-feb-2013	100.0

- What's an Entity Store?
- Existing Entity Stores
- Architecture
  - Catalog
  - Matching Engine
  - Facts
- Use Cases
- Building a global entity store (internet company view)
- Risks
- Next Steps

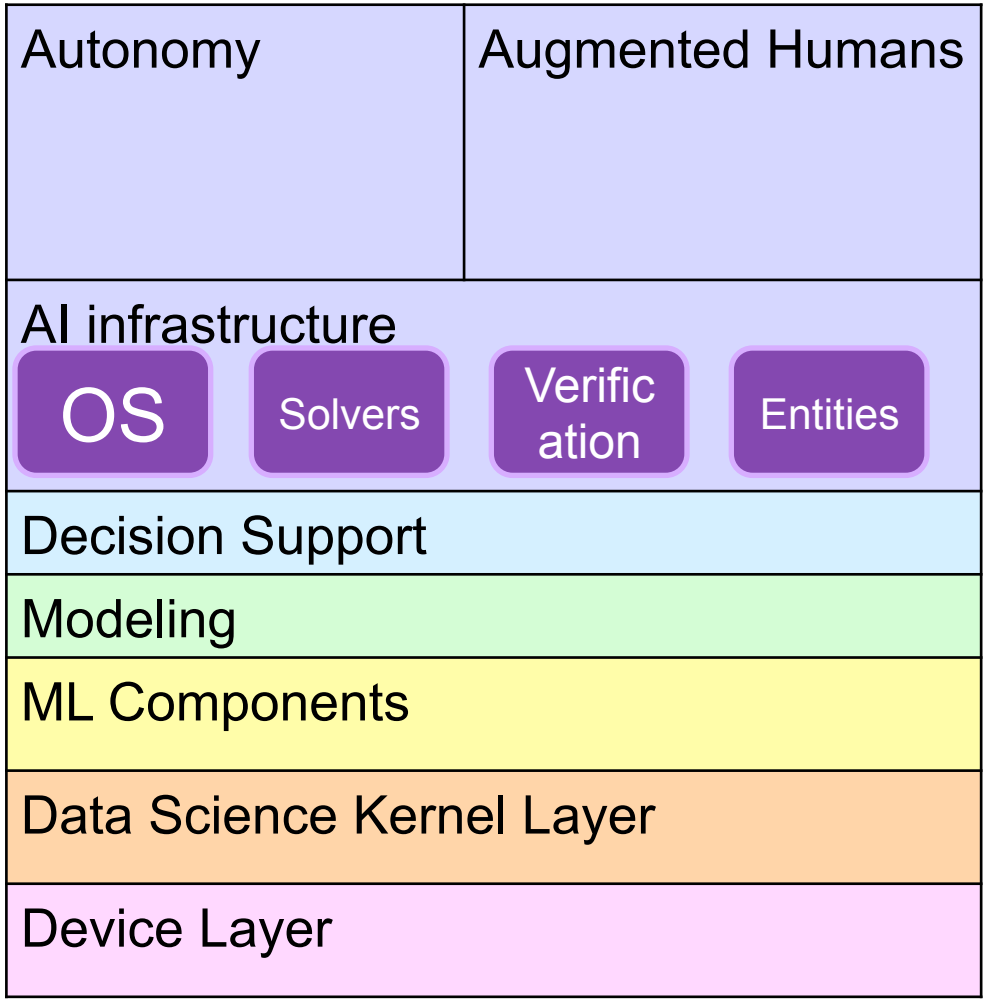
Technical Risks

- Undermerging, Overmerging, Multilevel taxonomies, Time, Uncertainty, Provenance.
- **Entity stores are alive:** You don't build an entity store once; you build a process to maintain, grow, and update a set of entities.
- **Physics and Sensing:** Many use cases (robotics and sensing) need to maintain information about visual, acoustic, and physics of physical-world objects.

Non-technical risks

- **Privacy.** Very serious problem. We recommend not including PII in such a project. There will need to be practical privacy technology in place to ask "what is the average age of women in Pittsburgh?" without having any explicit representation of all the people in Pittsburgh.
- **Provenance:** many major industries have their business model around obtaining facts.
- Why not leave this up to a large internet company to build? (Ans: this is bigger than Google or Apple or….)

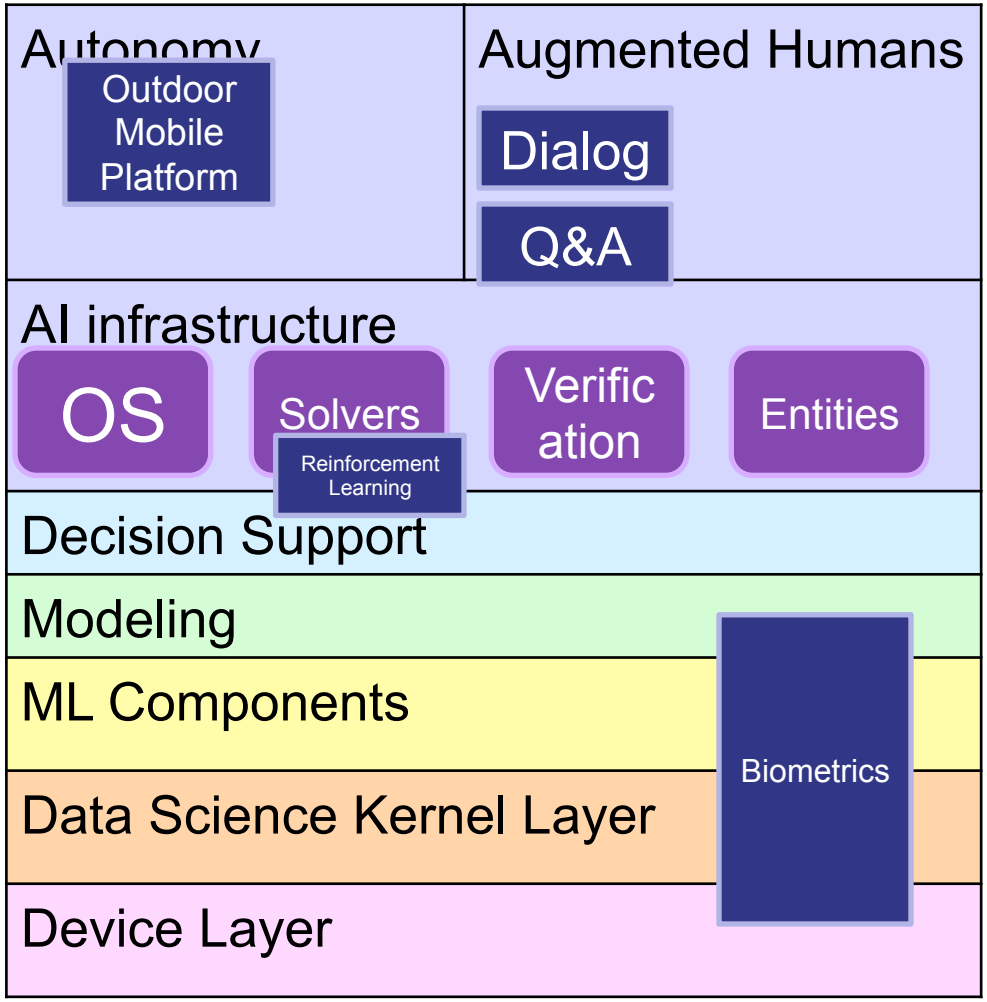
AI infrastructure



Industry focus areas

AI infrastructure component

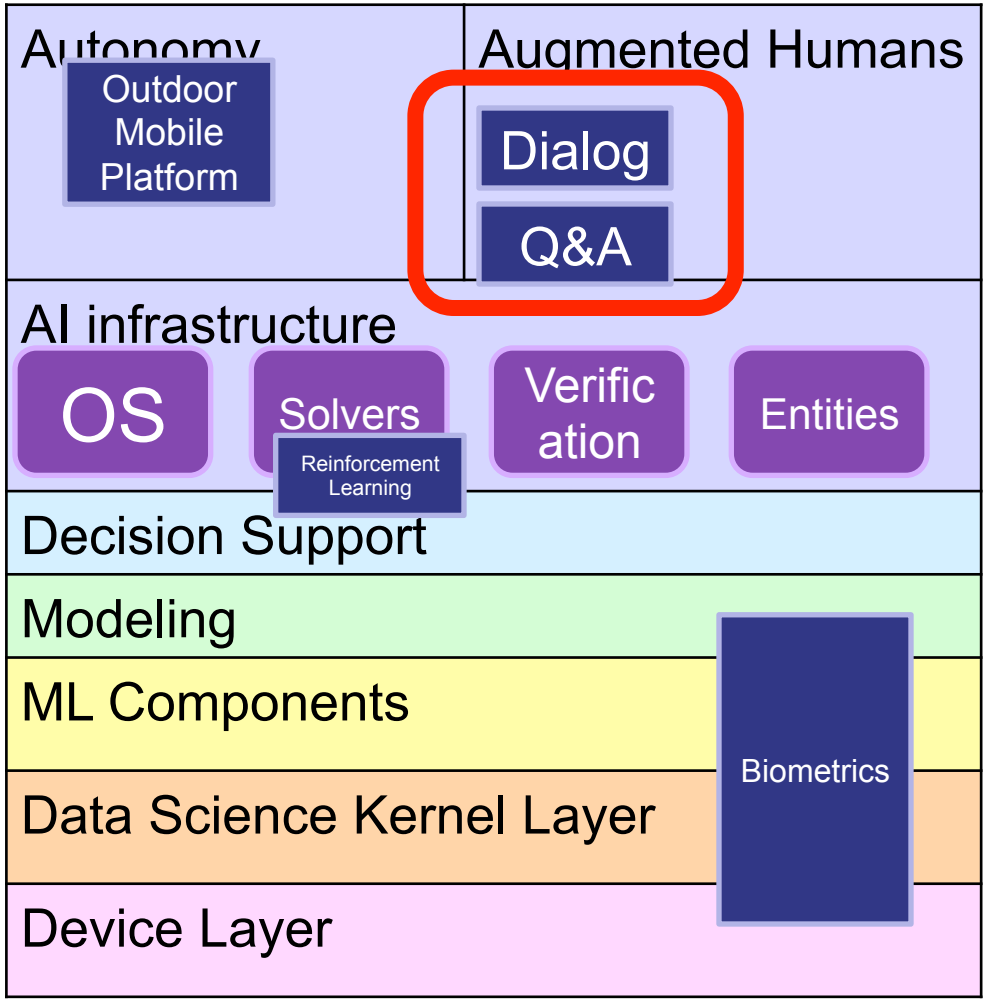
Industry Hotspot 2016



Industry focus areas

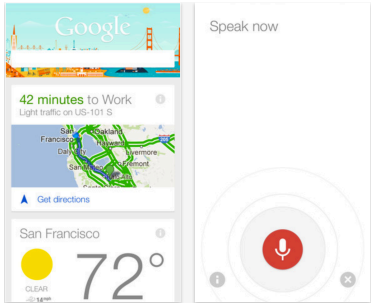
AI infrastructure component

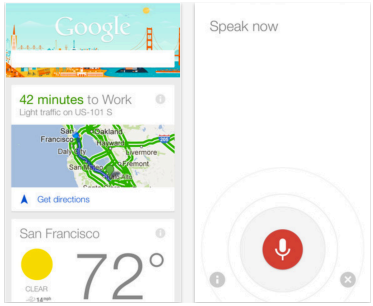
Industry Hotspot 2016

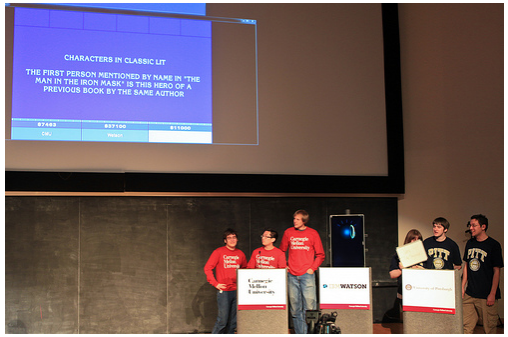



Siri


Microsoft
Hi. I'm Cortana.
Ask me a question!


Google
42 minutes to Work
Light traffic on US-101 S
Get directions
San Francisco
72°
Speak now





CMU/IBM
Entities and Question Answering

CMU/UPMC
Center for Machine Learning in Health

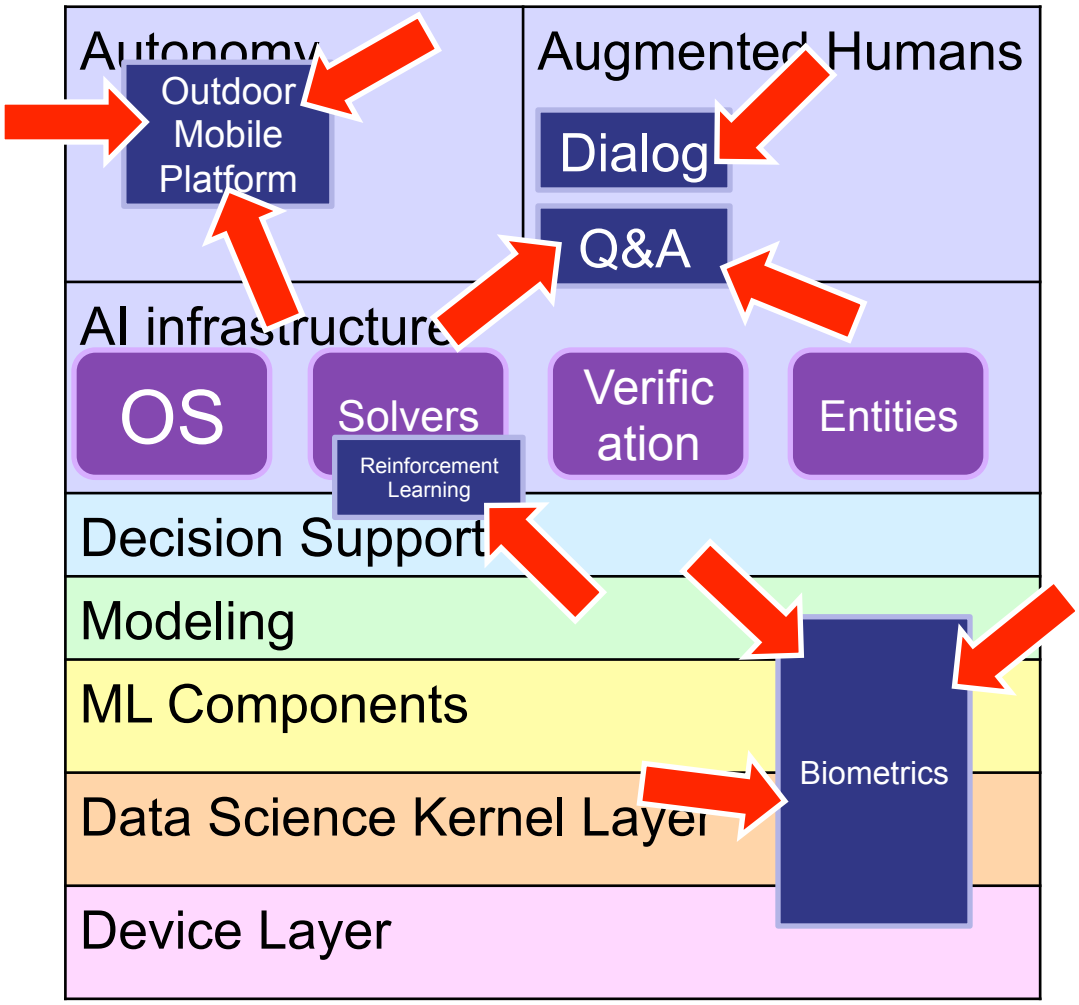
CMU/Boeing
Center for Aerospace Analytics

Carnegie Mellon University
School of Computer Science

Industry DANGER!!!

AI infrastructure component

Industry Hotspot 2016

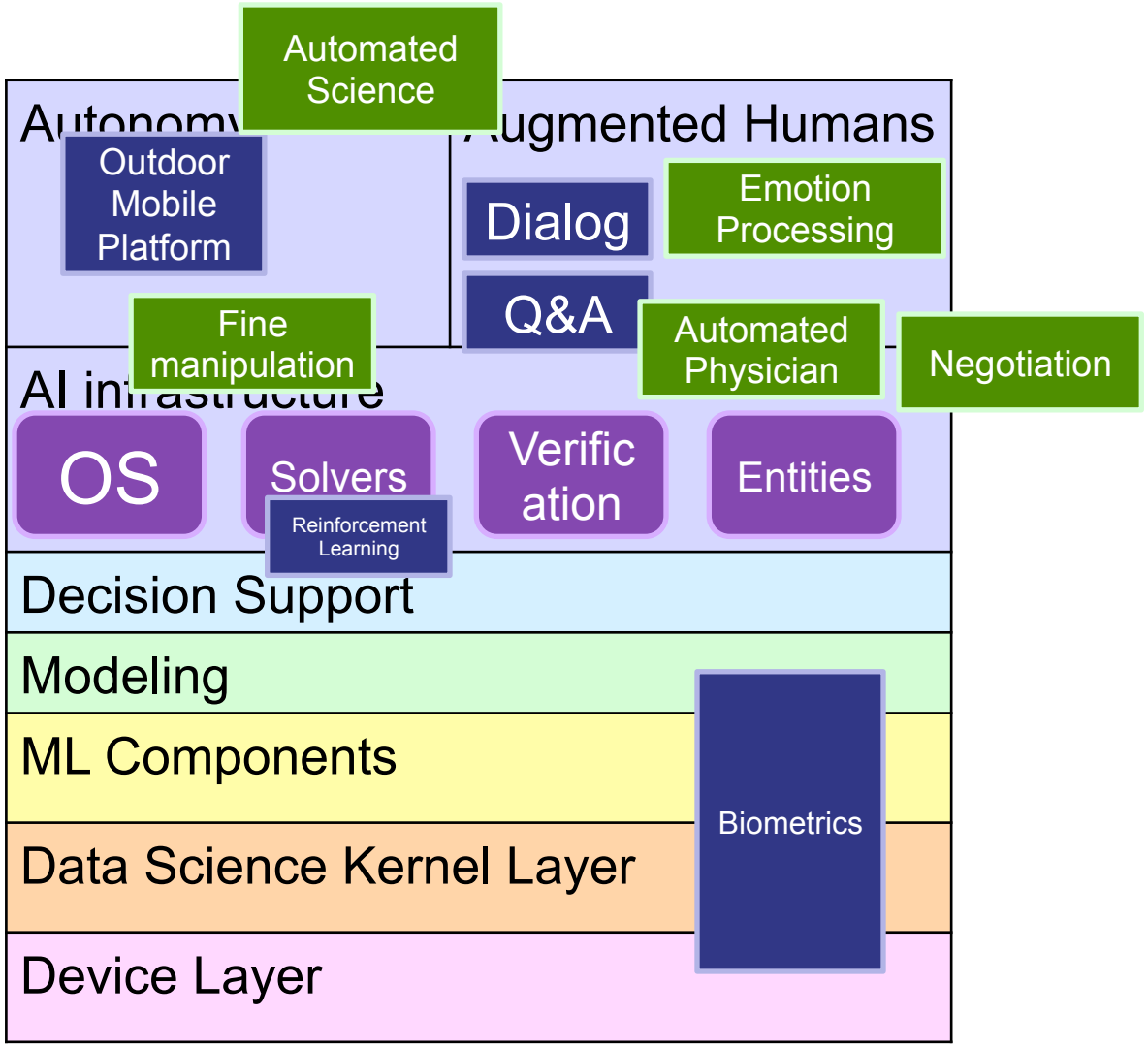


CMU focus areas

AI infrastructure component

Industry Hotspot 2016

CMU/SCS investment

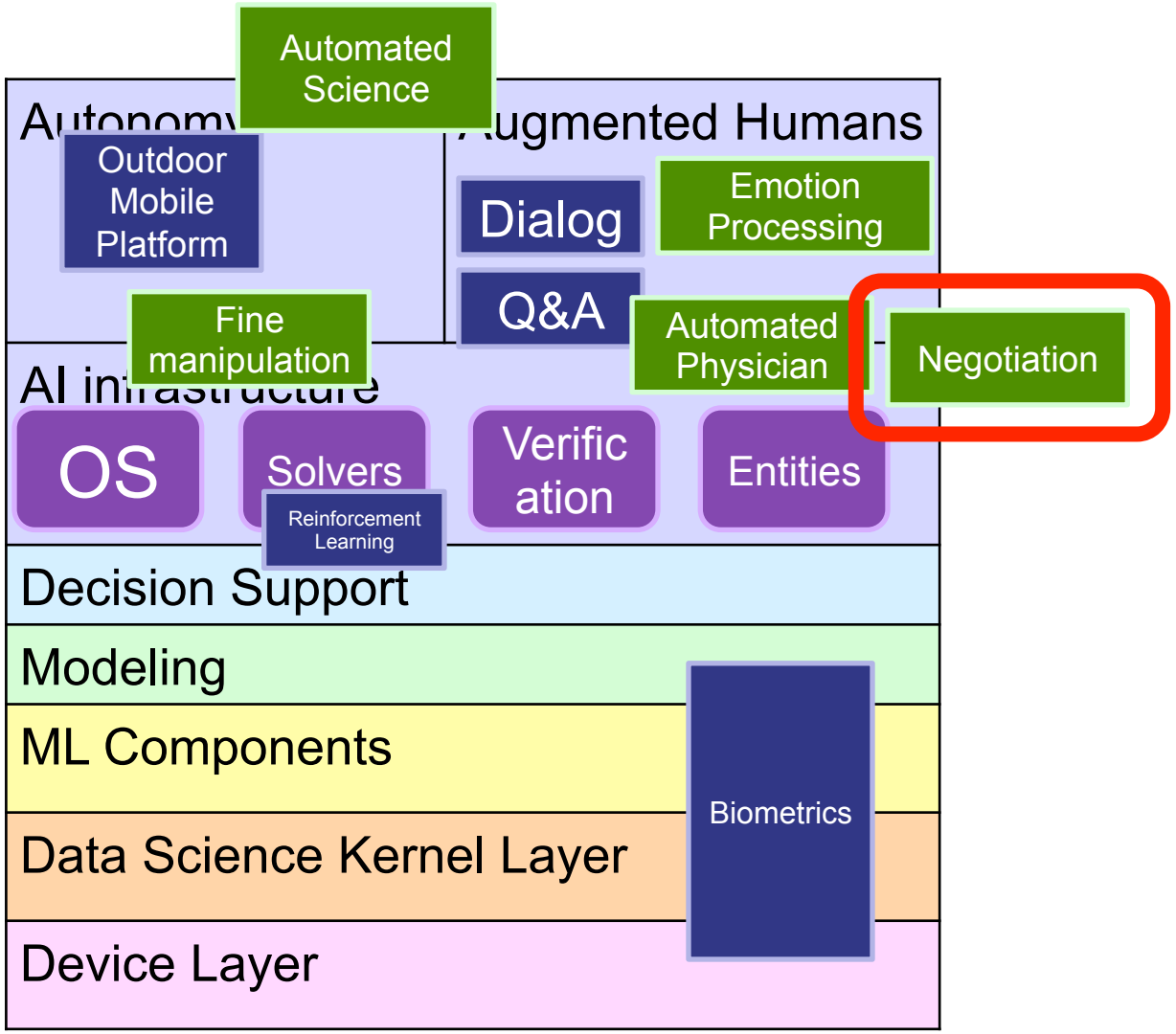


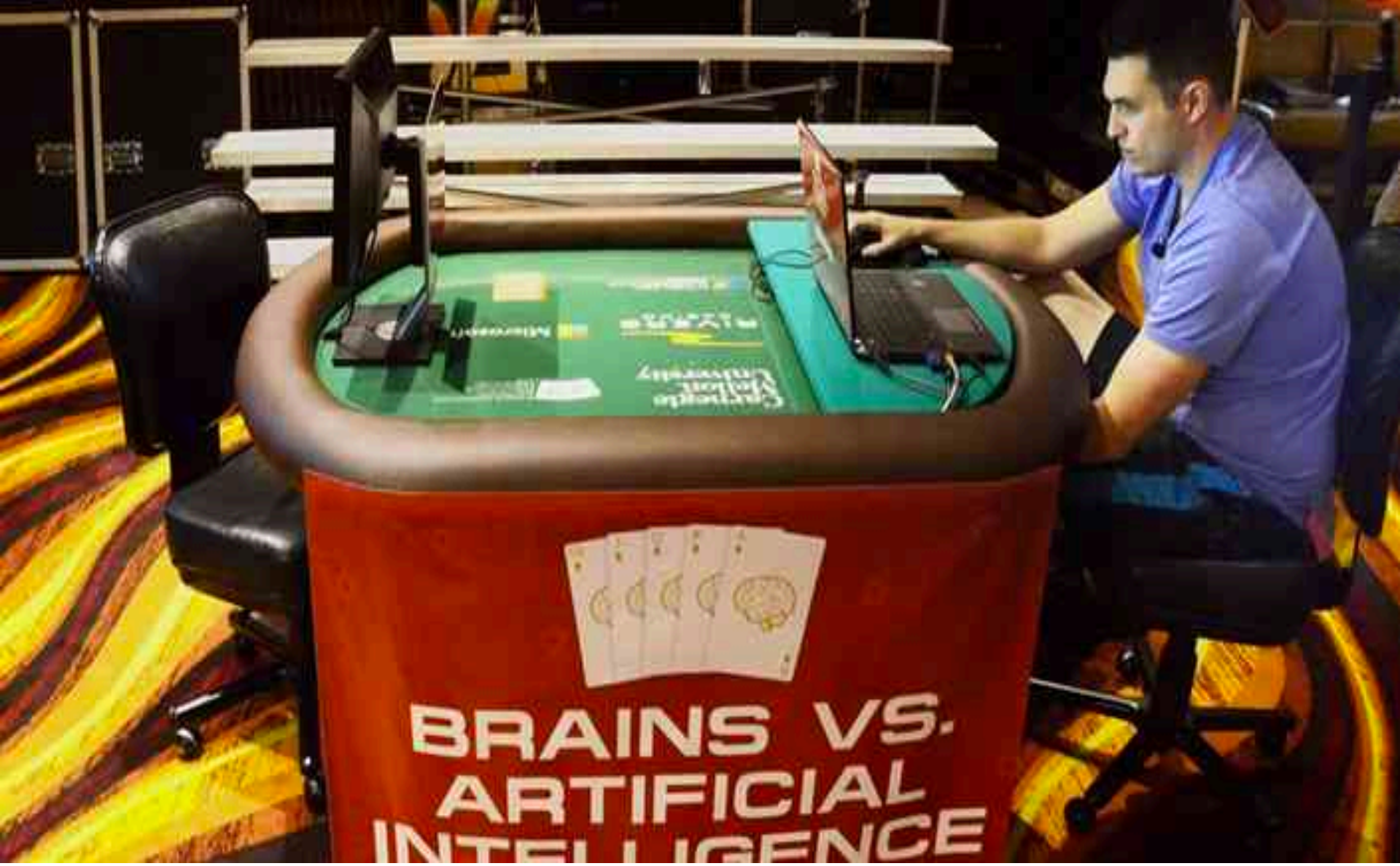
CMU focus areas

AI infrastructure component

Industry Hotspot 2016

CMU/SCS investment



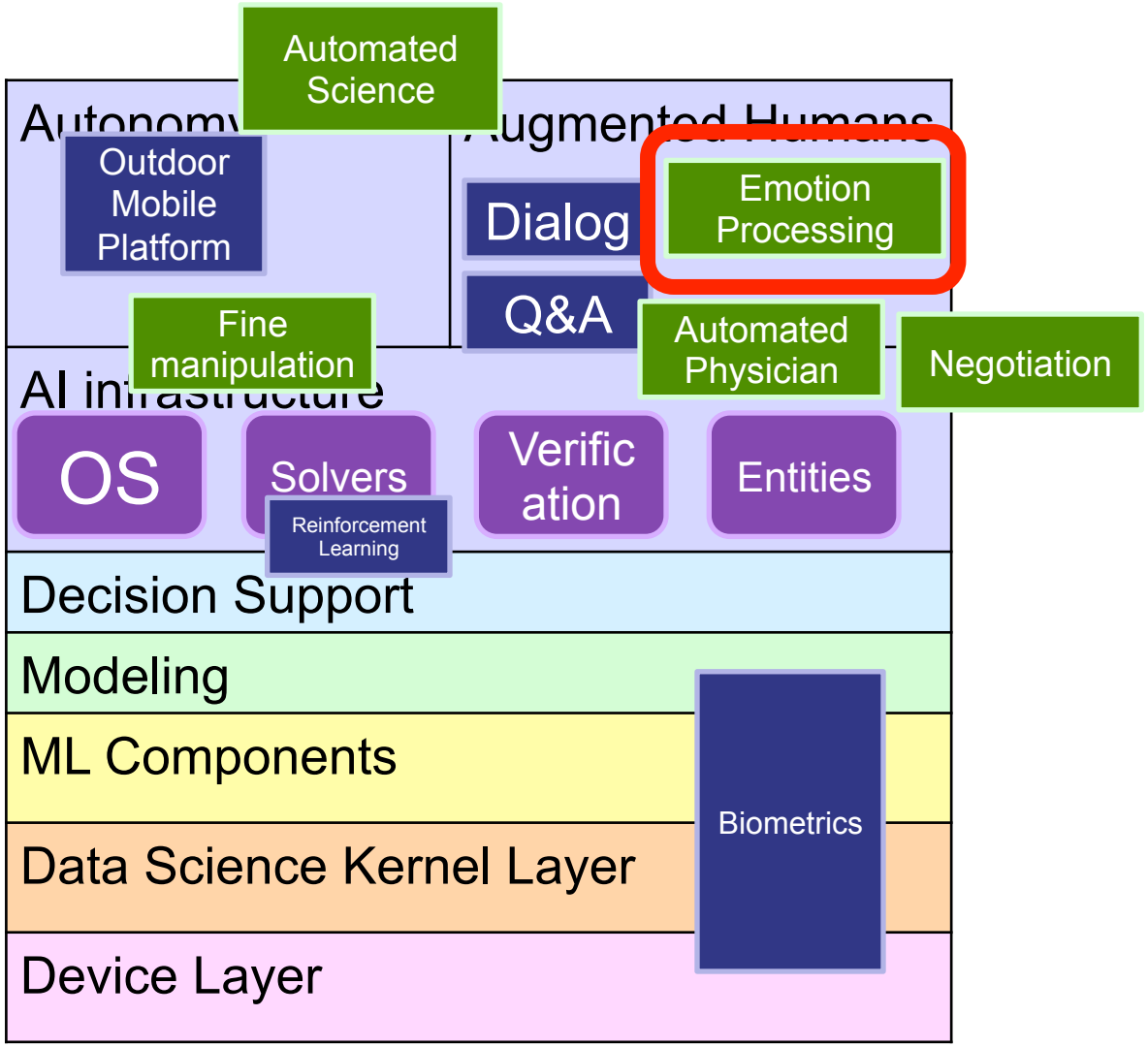

BRAINS VS.
ARTIFICIAL

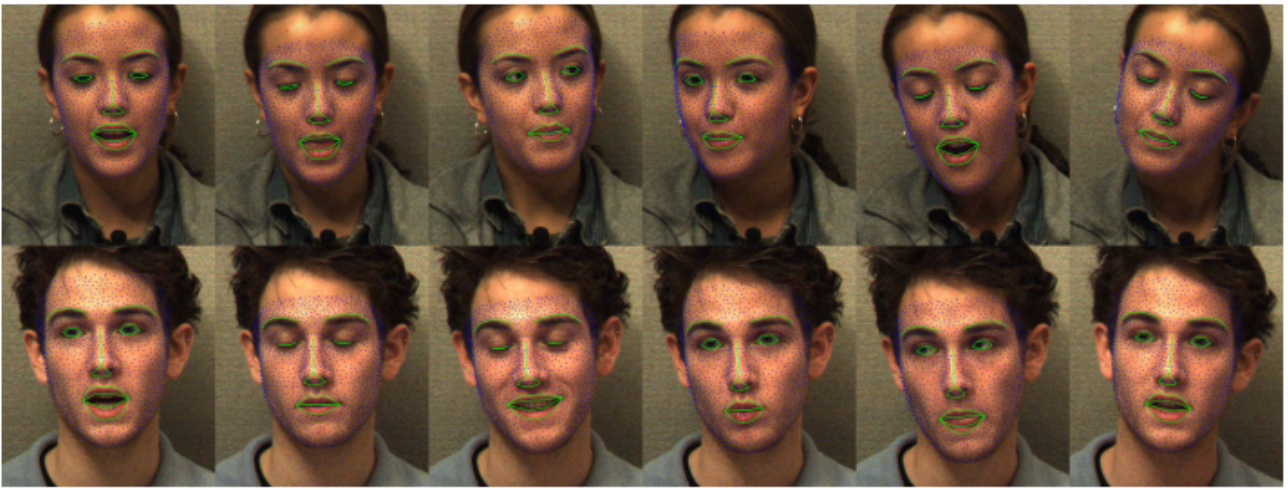
CMU focus areas

AI infrastructure component

Industry Hotspot 2016

CMU/SCS investment





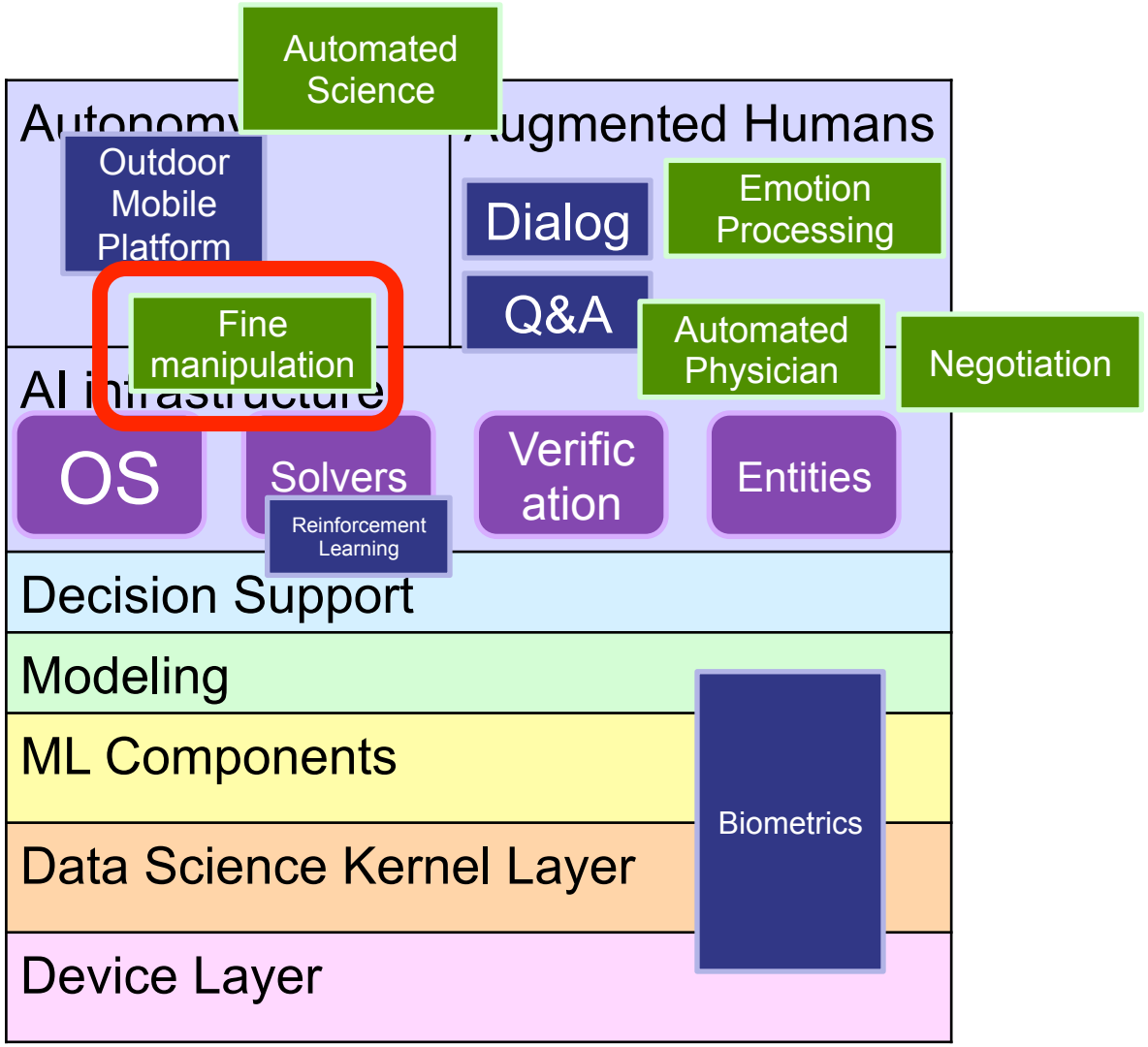
Understanding emotional features. Mental health metrics (Jeff Cohn (Pitt), Simon Lucey, LP Morency, Justine Casselle, Fernando de la Torre)

CMU focus areas

AI infrastructure component

Industry Hotspot 2016

CMU/SCS investment

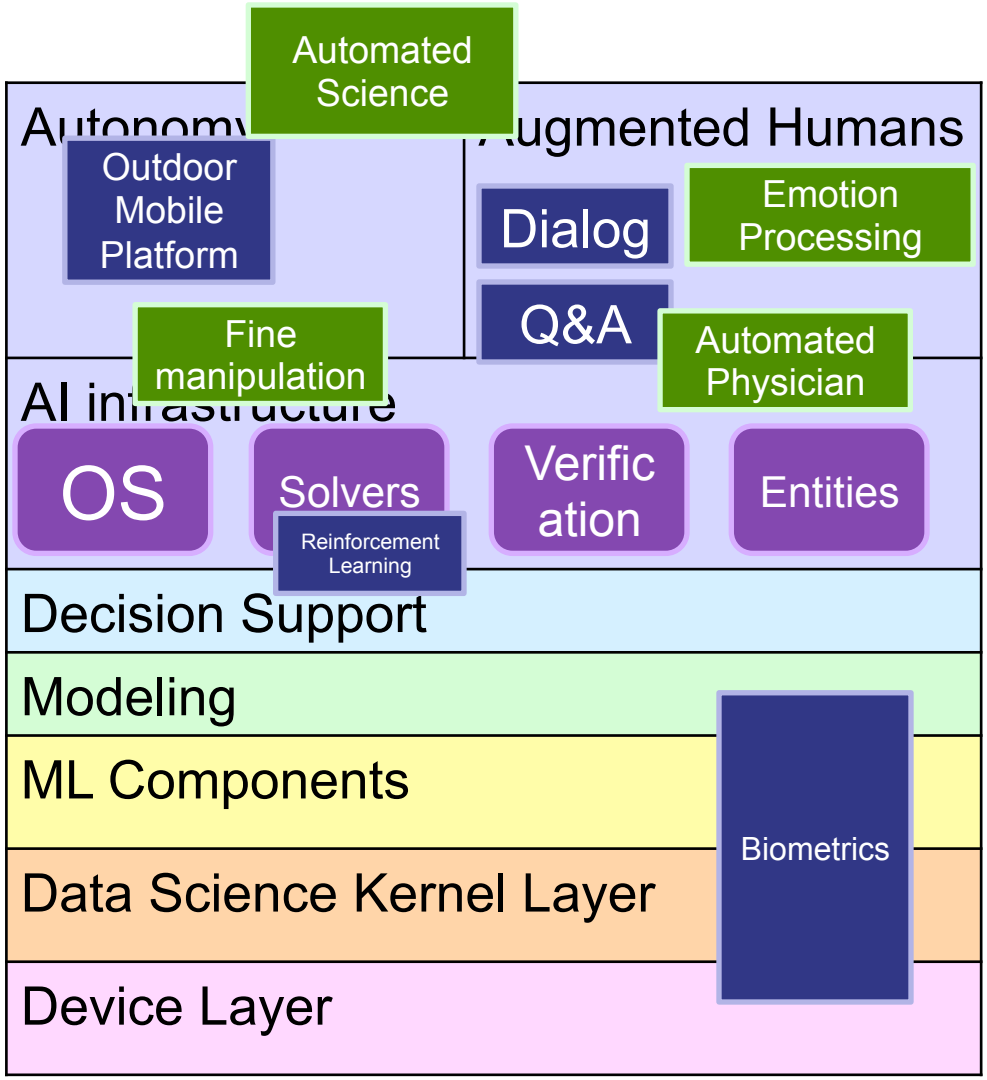


CMU focus areas

AI infrastructure component

Industry Hotspot 2016

CMU/SCS investment

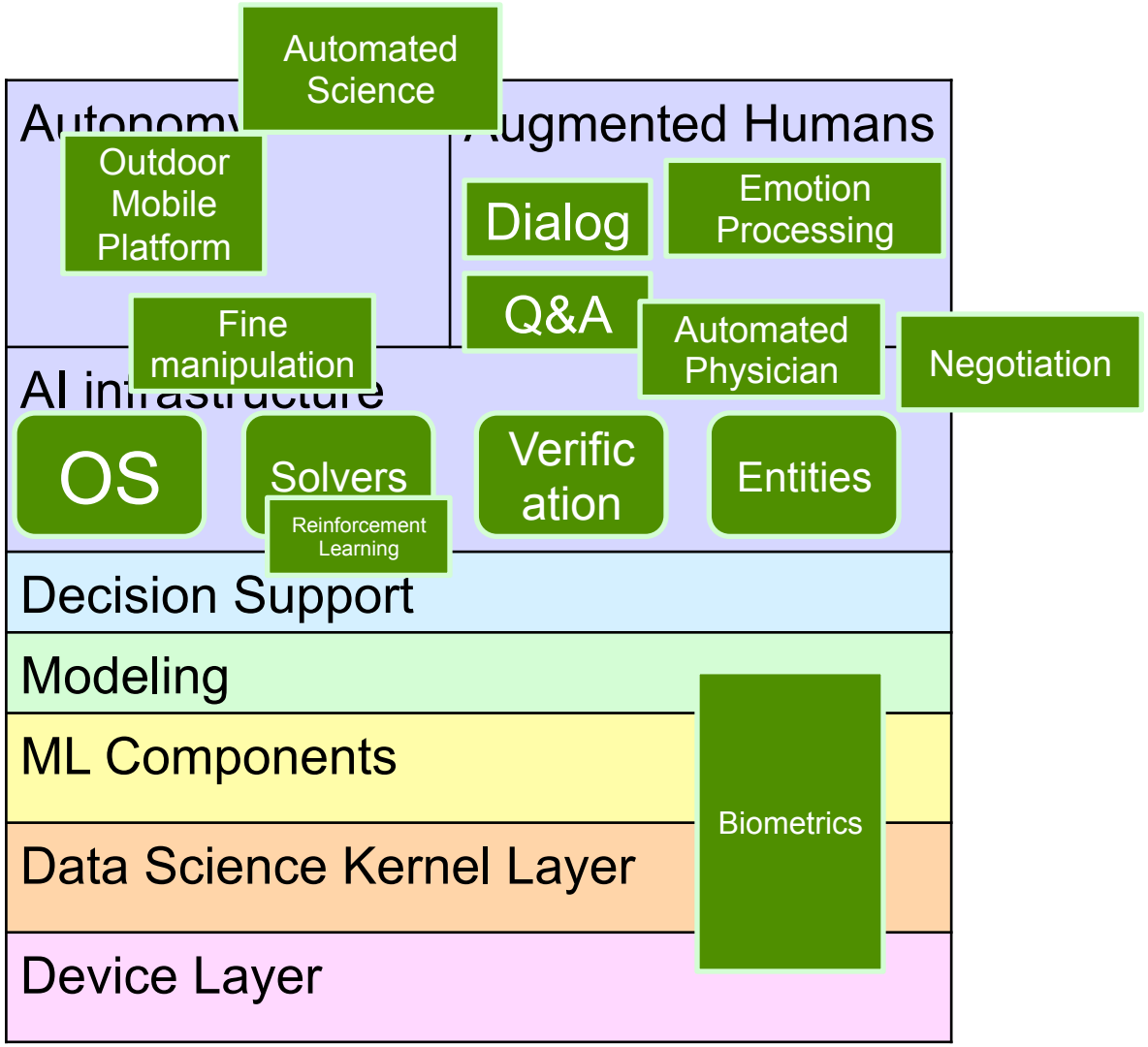


CMU focus areas

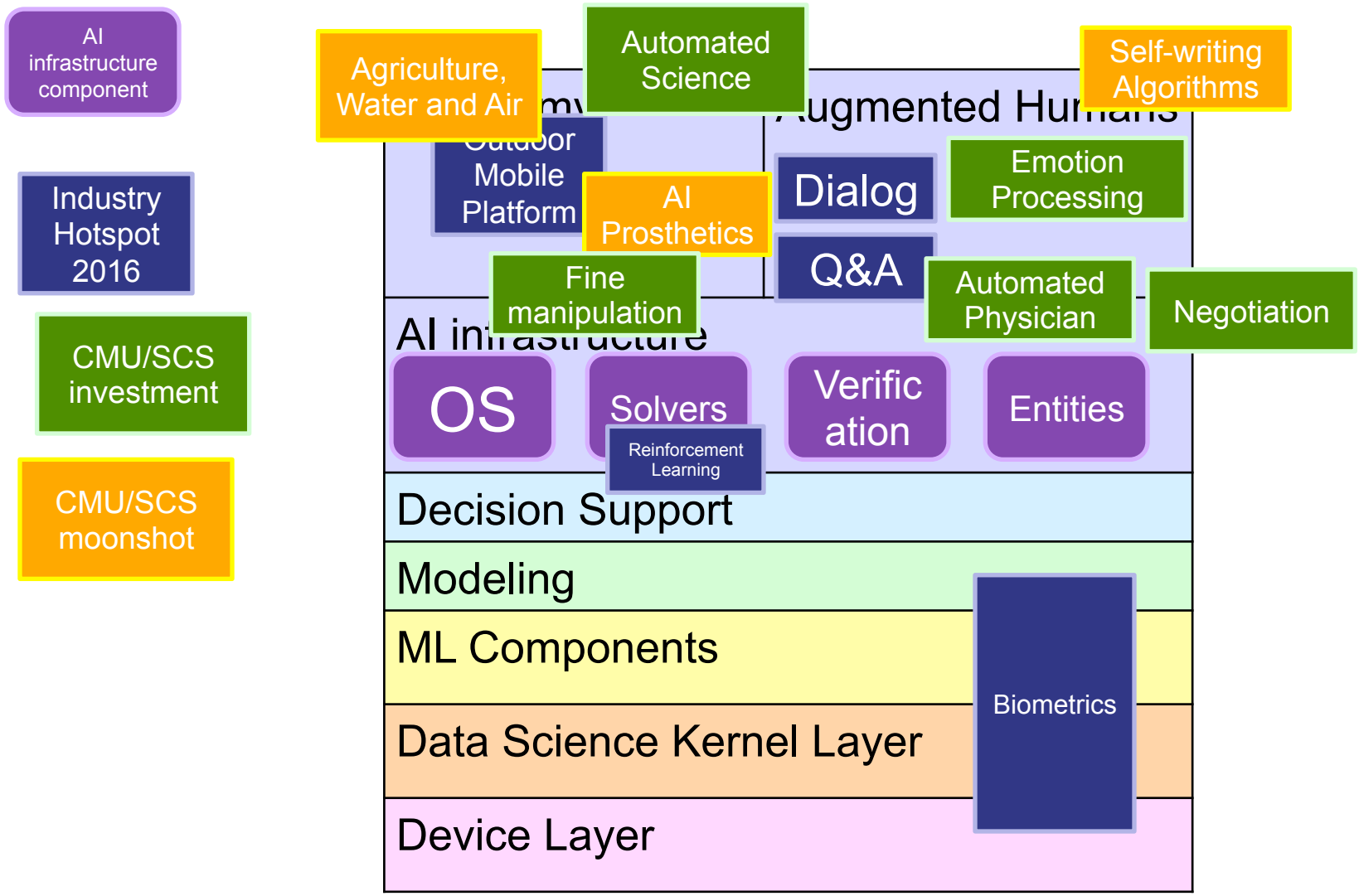
AI infrastructure component

Industry Hotspot 2016

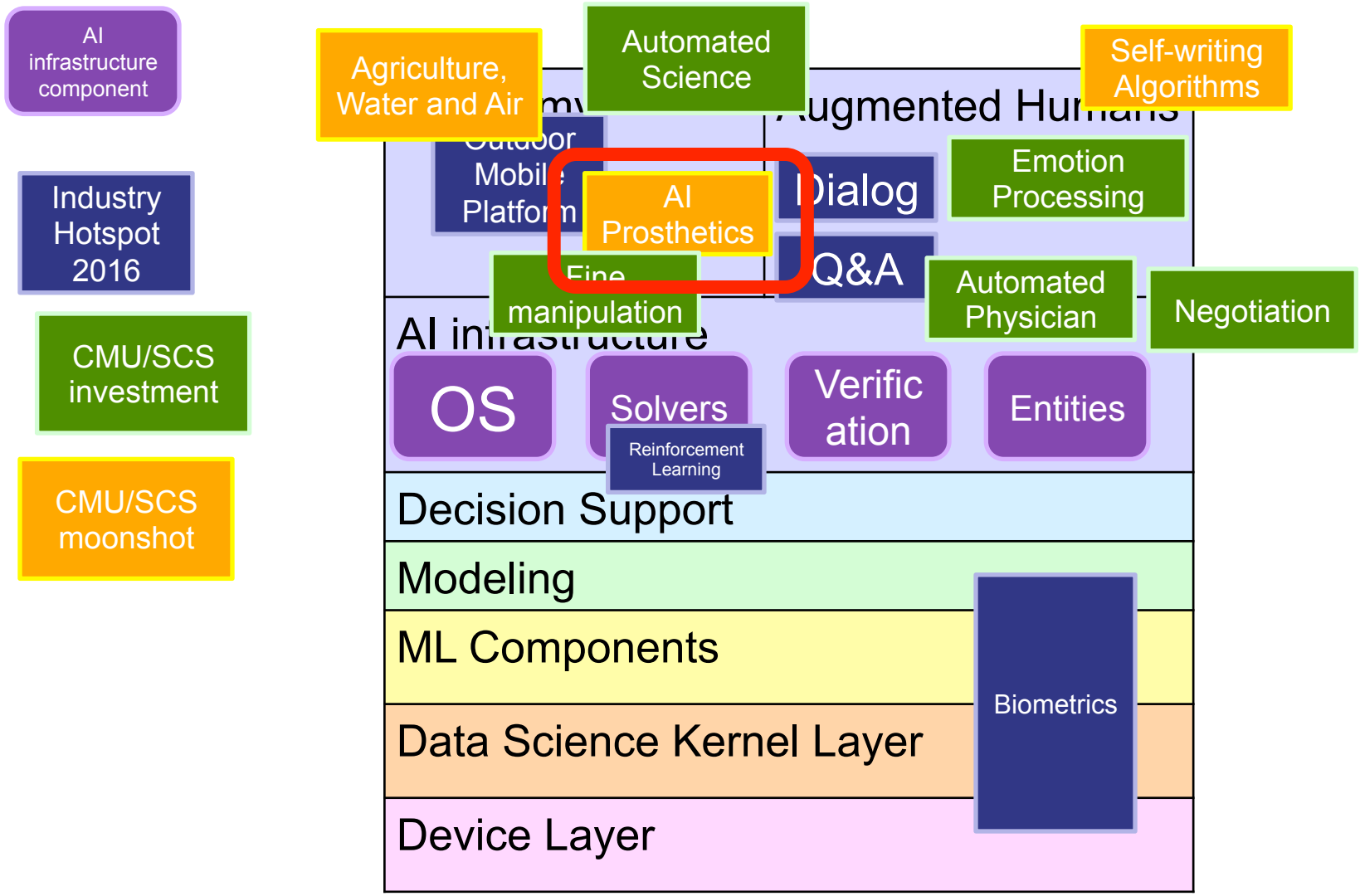
CMU/SCS investment

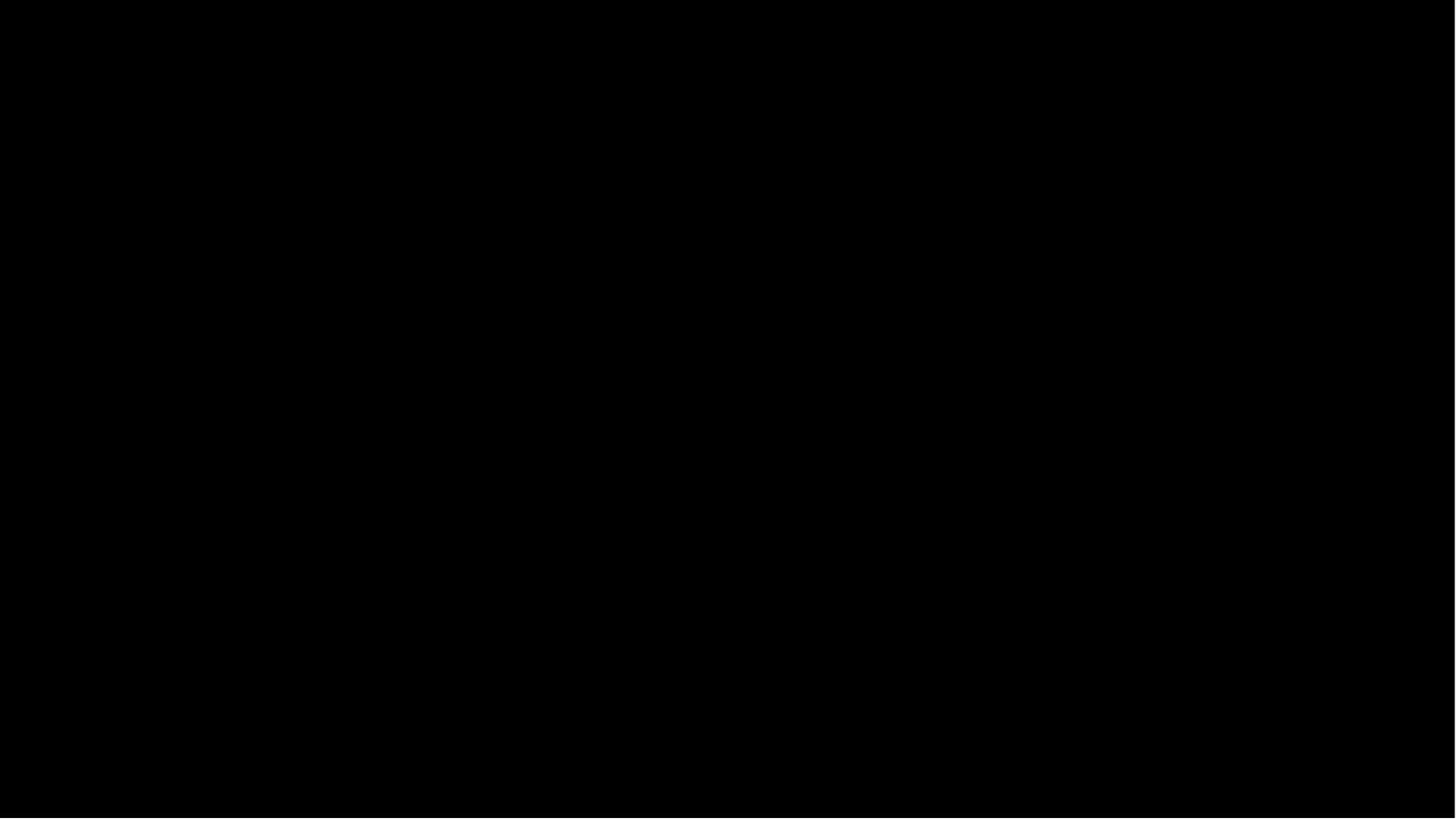


Moonshots

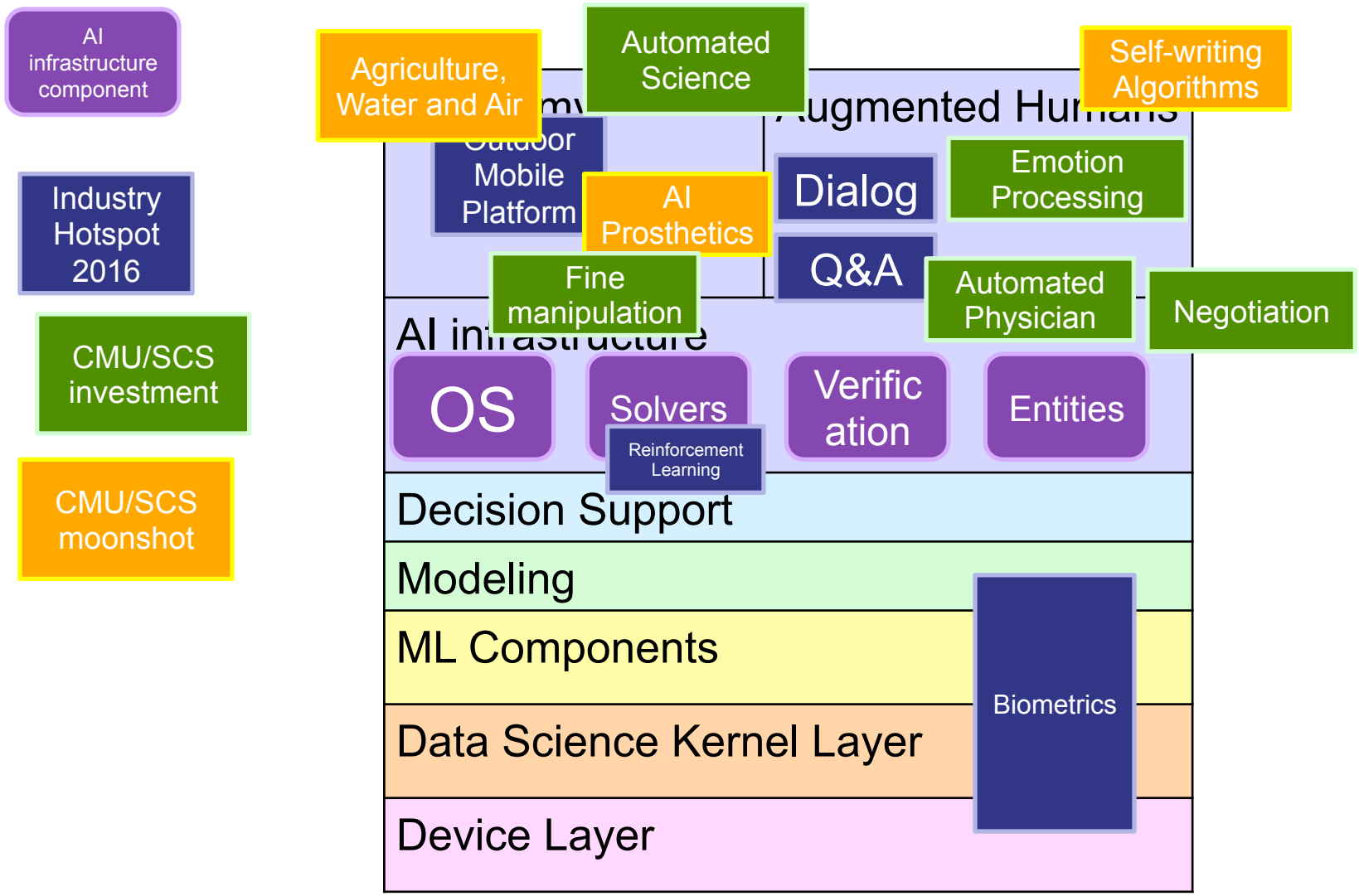

AI infrastructure component
Industry Hotspot 2016
CMU/SCS investment
CMU/SCS moonshot
Agriculture, Water and Air
Automated Science
Self-writing Algorithms
Outdoor Mobile Platform
AI Prosthetics
Dialog
Emotion Processing
Q&A
Fine manipulation
Automated Physician
Negotiation
AI infrastructure
OS
Solvers
Reinforcement Learning
Verification
Entities
Decision Support
Modeling
ML Components
Biometrics
Data Science Kernel Layer
Device Layer

Moonshots


AI infrastructure component
Industry Hotspot 2016
CMU/SCS investment
CMU/SCS moonshot
Agriculture, Water and Air
Automated Science
Self-writing Algorithms
Augmented Humans
Outdoor Mobile Platform
AI Prosthetics
Dialog
Emotion Processing
Q&A
Fine manipulation
Automated Physician
Negotiation
AI infrastructure
OS
Solvers
Reinforcement Learning
Verification
Entities
Decision Support
Modeling
ML Components
Biometrics
Data Science Kernel Layer
Device Layer



It's not just technology…



It's not just technology…