ihmc

# ihmc

## FLORIDA INSTITUTE FOR HUMAN & MACHINE COGNITION 🔍

Menu

- Research
  - Agile Computing
  - Autonomy & Coactive Design
  - Computational & Philosophical Foundations
  - Cybersecurity
  - Human Health, Resilience & Performance
  - Human-Machine Communication & Language Processing
  - Intentions, Beliefs & Trust
  - Knowledge Discovery, Data Science, Learning from Big Data
  - Knowledge Modeling, Work Analysis & Simulation, and Expertise Studies
  - Augmentics
  - Robotics, Exoskeletons, & Human Robotic Interdependence
  - Visualization & Human-Centered Displays
- CmapTools
- People
- About
  - The IHMC Story
  - Board of Directors
  - Science Advisory Council
  - Opportunities
  - Contact
  - IHMC Internal
- Outreach
  - Evening Lectures
  - Science Saturdays
  - Robotics Camp 2018
  - News
  - Newsletters
  - STEM-Talk

# Knowledge Discovery, Data Science, Learning from Big Data

Advances in technology have made it ever easier to collect and store data.  While the amount of data has increased tremendously, data collection is often opportunistic: There is only limited control over what, when, and how much data is collected, leading to inconsistency and fragmentary and uneven coverage. In addition, the connections one might seek from the data are almost always non-deterministic and uncertain. Traditional statistical and logical analyses are inadequate in many of these situations.

IHMC researchers James Allen, Greg Dubbin, Clark Glymour and Choh Man Teng work on a broad range of topics in Data Science, focusing on the development of methods to automatically analyze volumes of imperfect data to identify underlying connections. Although these researchers come from diverse backgrounds—linguistics, philosophy and computer science—the theme that unifies their research is the development of mechanisms that can learn from vast amounts of data and reason and understand the underlying structures that give rise to such data. Past and current projects include coastal zone mapping and target detection from LIDAR and sonar data, evidence-based paraconsistent reasoning, detecting and correcting data imperfections in satellite data, causal inference in climate teleconnections, forest fire predictions, and longitudinal modeling of glaucoma.

At the center of IHMC's work on data science is a knowledge base consisting of general knowledge about the world, instantial knowledge about the specific instances of interest, and statistical knowledge reflecting objective relative frequencies observed in the world.  The statistical statements are by nature approximate, and they can be represented by intervals (at a certain level of confidence) whose values and widths are indicative of the relative frequency and amount of supporting evidence.

Reference sets relevant to a particular event are constructed dynamically from the knowledge base, and principles to adjudicate between inconsistent statements are being developed to take into account such information as the taxonomy of the reference sets, the set inclusion relationship between the intervals, and background knowledge.  Similar principles are also used to identify and correct erroneous data, in a method developed at IHMC called *polishing*.

IHMC researchers have designed paradigms that transcend predictions and work on identifying the causal temporal connections between elements in the data.  This is important for generating an explanation of the constructed models and their predictions, in particular when the goal is to influence the outcome of an event of interest by influencing selected elements under (partial) control.  Even though both the causes and effects of a variable are correlated with the variable itself, one can only expect to influence the state of the variable by changing its causes but not its effects.

Algorithms are being developed that construct graphical causal models based on the conditional and unconditional probabilistic relations obtained in the data.  In the context of the DRUM: Deep Reader for Understanding Mechanisms project (funded by DARPA's "Big Mechanism" program), temporal and causal reasoning capabilities are being developed that can connect model fragments derived from both background knowledge and from reading scientific papers in English. Also in the area of Natural Language Processing, machine-learning methods are being developed to understand low resource languages, where syntactic and semantic information is scarce.
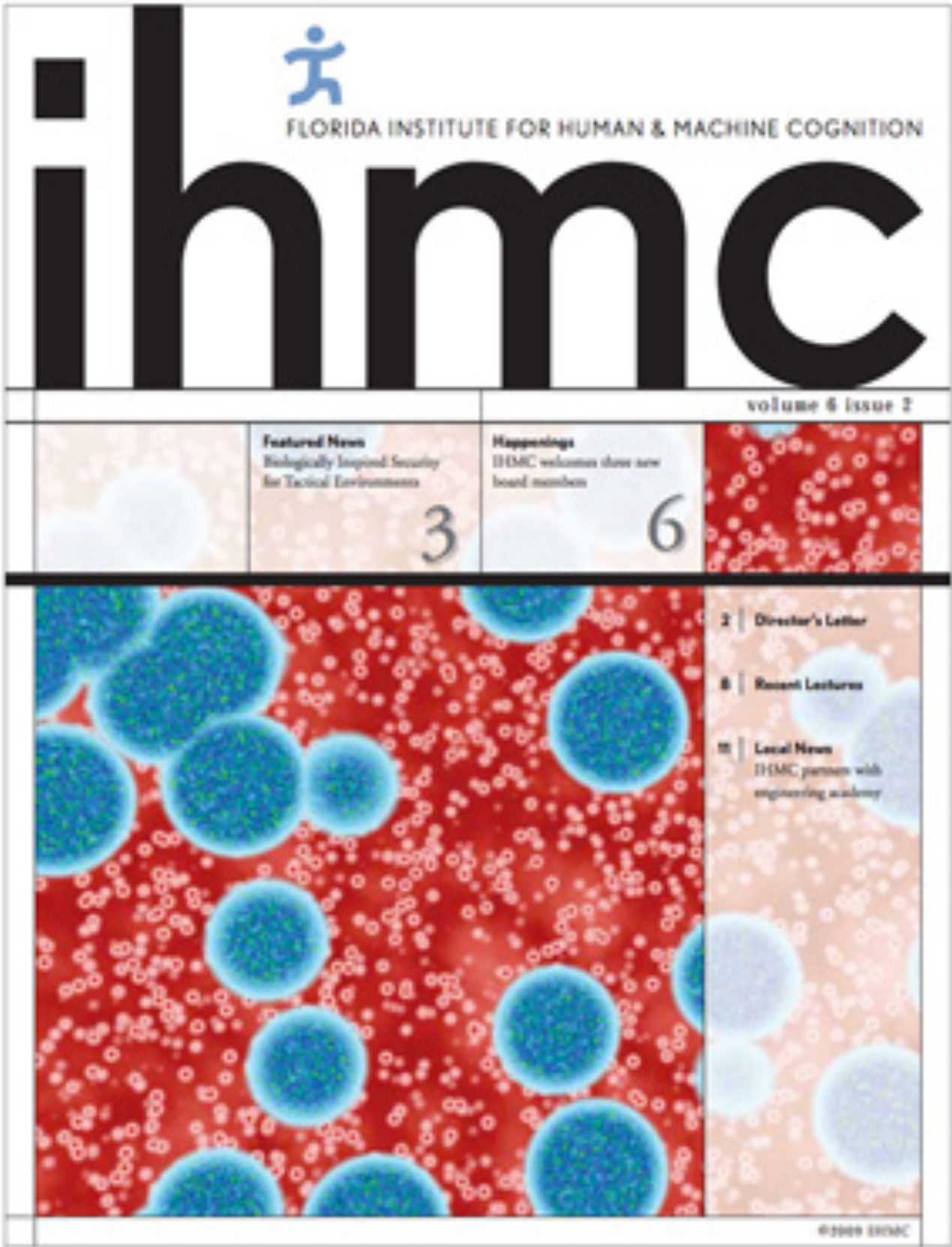
## Research Projects

- DRUM: Deep Reader for Understanding Mechanisms

## Research Team

- [Choh Man Teng](#)
- [James Allen](#)
- [Clark Glymour](#)

**Related Newsletters**

## FLORIDA INSTITUTE FOR HUMAN & MACHINE COGNITION

# ihmc

volume 6 issue 2

©2009 IHMC

FLORIDA INSTITUTE FOR HUMAN & MACHINE COGNITION

# ihmc

VOLUME 10 ISSUE 1

**Featured Research**
New cyber threats, new security tools
**3**

**New Research**
Computers developing understanding of language
**7**

**Happenings**
IHMC-NASA partnership on Earth and in space
**10**

**Happenings**
Cmap conference lands in Malta
**12**

©2015 IHMC

Researchers at IHMC pioneer technologies aimed at leveraging and extending human capabilities. Our human-centered approach often results in systems that can be regarded as cognitive, physical, or perceptual orthoses, much as eyeglasses are a kind of ocular orthoses. These systems fit the human and machine components together in ways that exploit their respective strengths and mitigate their respective weaknesses. Read More »

## Latest News

- Levin Center for IHMC Research wins AIA Florida's People's Choice Award
- Kernagis wins Outstanding Young Scientist Award
- Levin Center for IHMC Research nominated for architectural award
- Ewers donate $500,000 to IHMC
- Studers Donate $1 Million to Further IHMC New Initiatives

## Newsletters

**Pensacola Campus**

40 South Alcaniz St.
Pensacola, FL 32502
850.202.4462
e-mail
Google Maps

**Ocala Campus**

15 SE Osceola Ave.
Ocala, FL 34471
352.387.3050
e-mail
Google Maps

Privacy Policy

© 2018 IHMC