SĀF ART INT

WORKSHOP ON
SAFETY & CONTROL FOR
ARTIFICIAL INTELLIGENCE

Models and validation

From Mike Wagner
presentation at
SAF|ART|INT

Photo: driversed.com

"The car shall [stop] for [pedestrians] [in a crosswalk]."

Models and validation

From Mike Wagner presentation at SAF|ART|INT

Models and validation

From Mike Wagner presentation at SAF|ART|INT

Models and validation

"**Validation:** The process of determining that the requirements are the correct requirements and that they are complete."

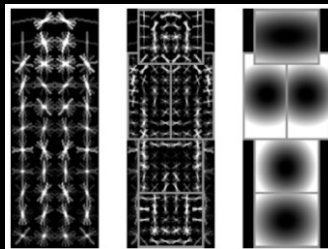Software Considerations in Airborne Systems
and Equipment Certification
(RTCA DO-178C)

Models and validation

and…   so: A pedestrian is present

Conclusion

Because we observe [**data**], and since we agree that [**warrant**], then you should accept that [**conclusion**] is true.

Toulmin Model of Argument

Ethical dilemmas

From Andrew Moore presentation at SAF|ART|INT

Ethical dilemmas

From Andrew Moore presentation at SAF|ART|INT

1 in 10 million

75%

1 in a million

1%

June 28, 2016 – White House OSTP and Carnegie Mellon Workshop

| Introduction	Ed Felten (White House OSTP) Bill Scherlis (CMU)
AI applications, safety requirements, big ideas	Manuela Veloso (CMU) John Launchbury (DARPA) Jason Matheny (IARPA) Andy Grotto (White House NSC)
AI Algorithms, Intrinsic Safety, Explanations, Evaluations	Claire Tomlin (UC Berkeley) Tom Dietterich (Oregon State) Emma Brunskill (CMU) Michael Littman (Brown)
Mathematical Models and Reasoning	Jeannette Wing (Microsoft) Kathleen Fisher (Tufts) Anupam Datta (CMU) Sarah Loos (Google)
Systems and Safety Engineering	Mike Wagner (CMU) Drew Bagnell (Uber) Reid Simmons (NSF) Brian Murray (ZF) Doug Schmidt (Vanderbilt)

June 27, 2016 – Carnegie Mellon Exploratory Workshop

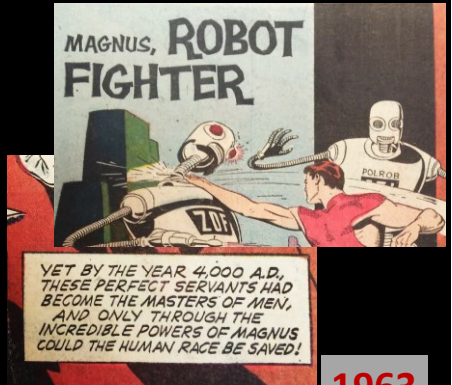| Context	Ed Felten (White House OSTP) Bill Scherlis (CMU) Eric Horvitz (Microsoft)
Perspective	Andrew Moore (CMU) Richard Mallah (Future of Life Inst) Tom Mitchell (CMU
Focus	Dario Amodei (Google Brain) Claire Le Goues (CMU) Robert Rahmer (IARPA)

1920
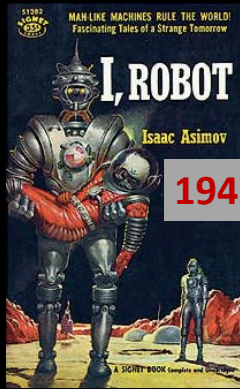
1942

MAGNUS, ROBOT FIGHTER

1963

1590ca