Trusted AI and AI Trust

An Opportunity for Synthesis

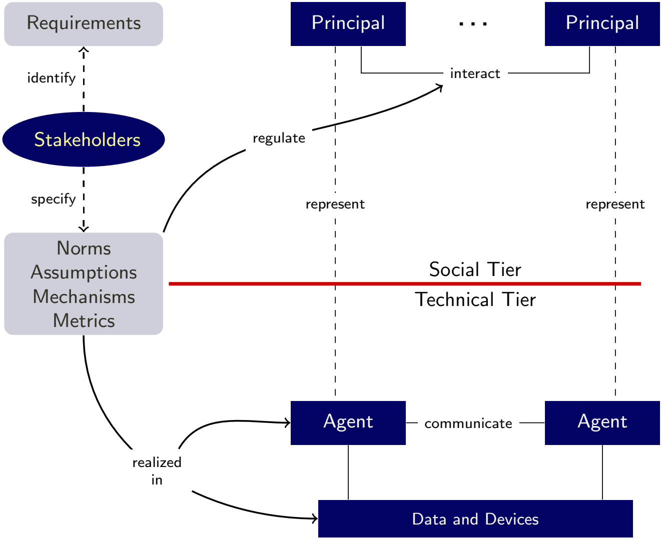
Munindar P. Singh
singh@ncsu.edu

Department of Computer Science
North Carolina State University

July 2018

Sociotechnical Systems

Current AI research: atomistic, single-agent decision-making and ethical dilemmas
Current social sciences research: Not computational in outlook



Comparison

Requirements

Scope

Aspect

Autonomy

Nature

Fairness

Research Focus

Comparison

	Trusted AI
Requirements	Of agents to people
Scope	Trustworthiness
Aspect	Instrumental: agents are tools
Autonomy	Intelligence and complexity
Nature	Transparency
Fairness	Statistical wrt protected groups
Research Focus	Individual dilemmas

Comparison

	Trusted AI	AI Trust
Requirements	Of agents to people	By agents of others
Scope	Trustworthiness	Trust
Aspect	Instrumental: agents are tools	Sociocognitive: agents are socially intelligent
Autonomy	Intelligence and complexity	Decision making wrt social relationships
Nature	Transparency	Accountability
Fairness	Statistical wrt protected groups	Individual wrt vulnerability
Research Focus	Individual dilemmas	Systemic properties

Conclusion

Going back to sociotechnical systems

- Build on sociocognitive modeling
- Incorporate human considerations of interpretability and understanding
- Incorporate reasoning about incentives
- Support composition
  - Systems of systems with . . .
  - Systems that appear as agents
  - Systems that appear as tools
- Toward a theory of ethics