

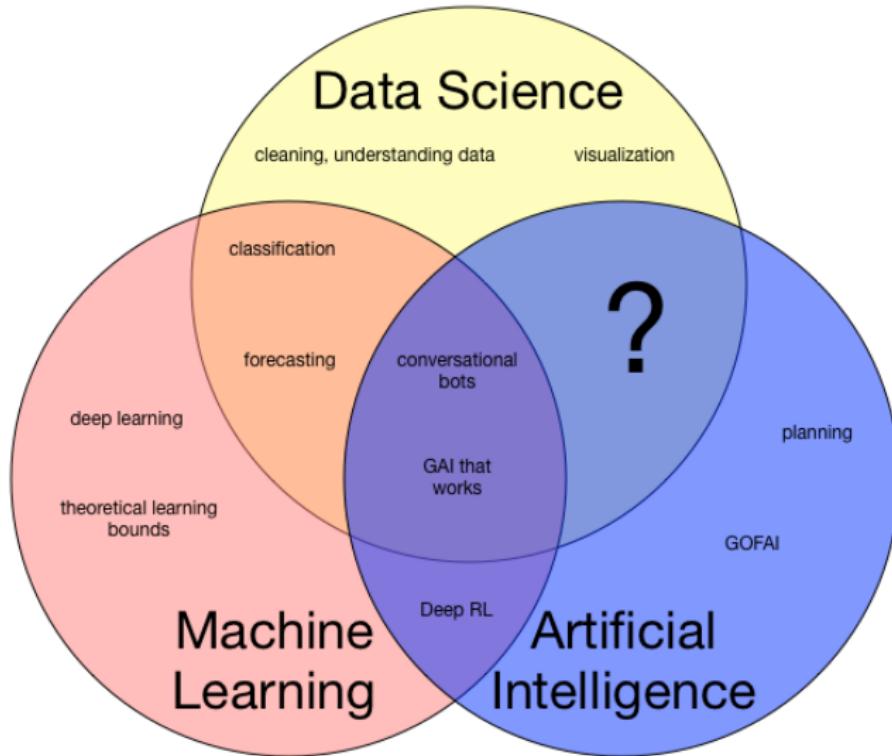


Definitions

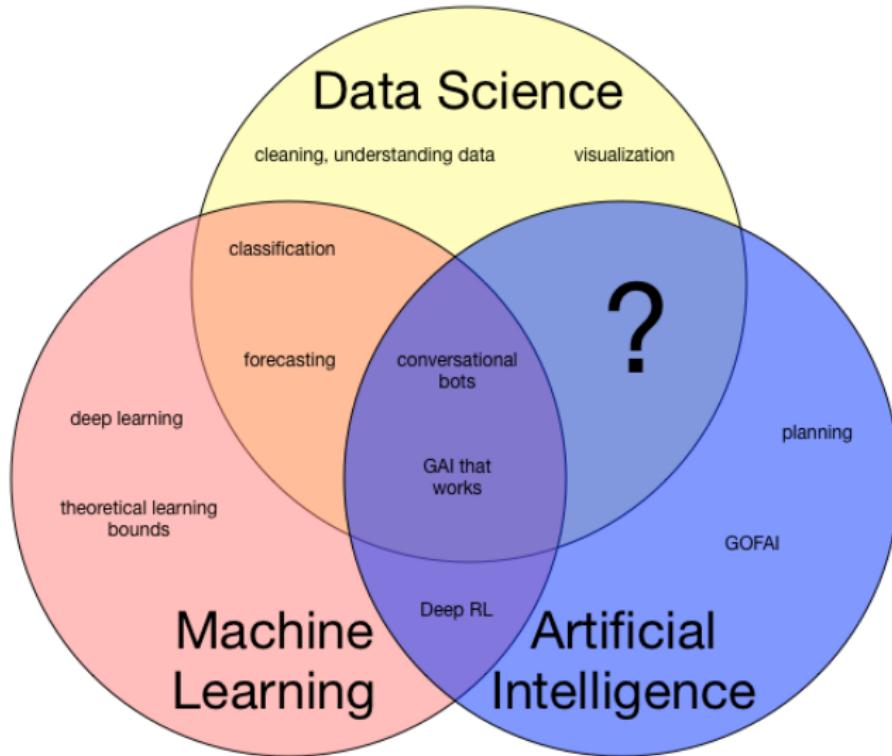
Jordan Boyd-Graber
University of Maryland
JANUARY 12, 2018

Roadmap

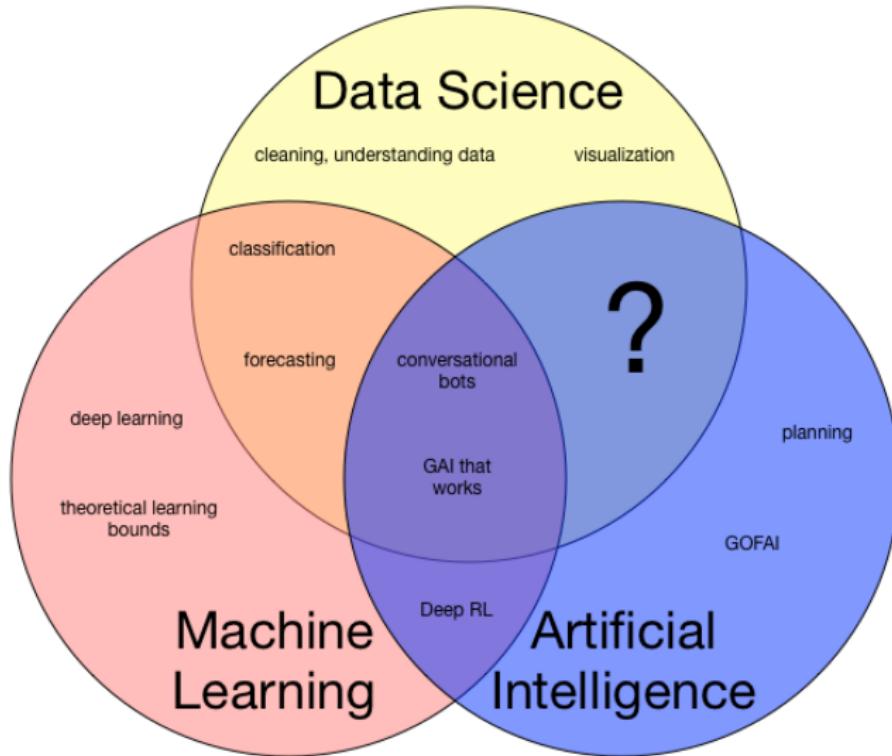
- What is data science / artificial intelligence / machine learning?
- What is possible with these new tools and techniques?
- What are the challenges?



Data provides the foundation



Machine learning provides algorithms



Artificial intelligence defines problems



Nowhere near true AI

Data are everywhere.

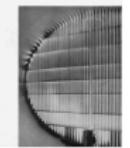
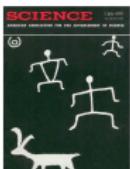
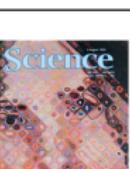
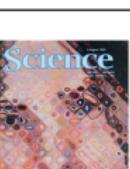
User ratings

<u>Ikiru</u> (1952)	UR	Foreign	
<u>Junebug</u> (2005)	R	Independent	
<u>La Cage aux Folles</u> (1979)	R	Comedy	
<u>The Life Aquatic with Steve Zissou</u> (2004)	R	Comedy	
<u>Lock, Stock and Two Smoking Barrels</u> (1998)	R	Action & Adventure	
<u>Lost in Translation</u> (2003)	R	Drama	
<u>Love and Death</u> (1975)	PG	Comedy	
<u>The Manchurian Candidate</u> (1962)	PG-13	Classics	
<u>Memento</u> (2000)	R	Thrillers	
<u>Midnight Cowboy</u> (1969)	R	Classics	

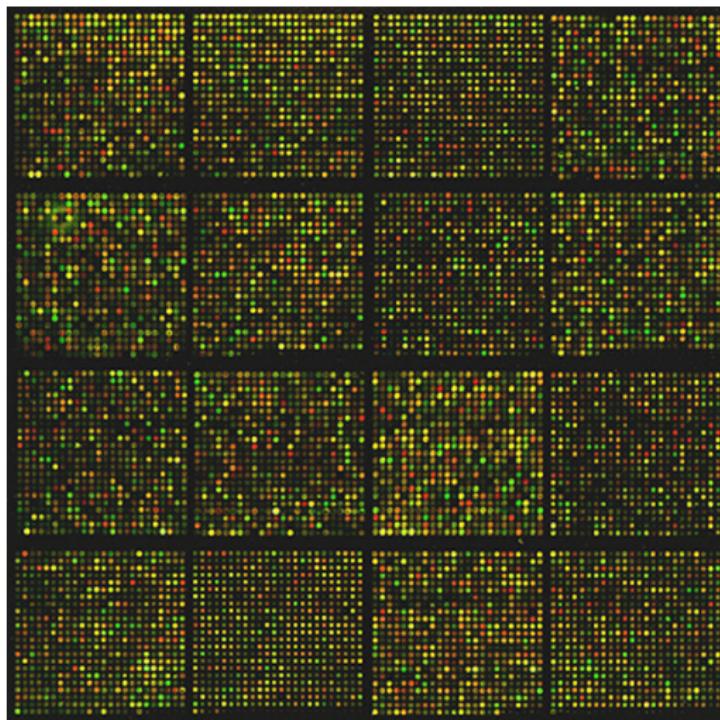
Purchase histories

Cheese				
0.5/0.51 lb	Cabot Vermont Cheddar	0.51 lb	\$7.99/lb	\$4.07
Dairy				
1/1	Friendship Lowfat Cottage Cheese (16oz)		\$2.89/ea	\$2.89
1/1	Nature's Yoke Grade A Jumbo Brown Eggs (1 dozen)		\$1.49/ea	\$1.49
1/1	Santa Barbara Hot Salsa, Fresh (16oz)		\$2.69/ea	\$2.69
1/1	Stonyfield Farm Organic Lowfat Plain Yogurt (32oz)		\$3.59/ea	\$3.59
Fruit				
3/3	Anjou Pears (Farm Fresh, Med)	1.76 lb	\$2.49/lb	\$4.38
2/2	Cantaloupe (Farm Fresh, Med)		\$2.00/ea	\$4.00 S
Grocery				
1/1	Fantastic World Foods Organic Whole Wheat Couscous (12oz)		\$1.99/ea	\$1.99
1/1	Garden of Eatin' Blue Corn Chips (9oz)		\$2.49/ea	\$2.49
1/1	Goya Low Sodium Chickpeas (15.5oz)		\$0.89/ea	\$0.89
2/2	Marcal 2-Ply Paper Towels, 90ct (1ea)		\$1.09/ea	\$2.18 T
1/1	Muir Glen Organic Tomato Paste (6oz)		\$0.99/ea	\$0.99
1/1	Starkist Solid White Albacore Tuna in Spring Water (6oz)		\$1.89/ea	\$1.89

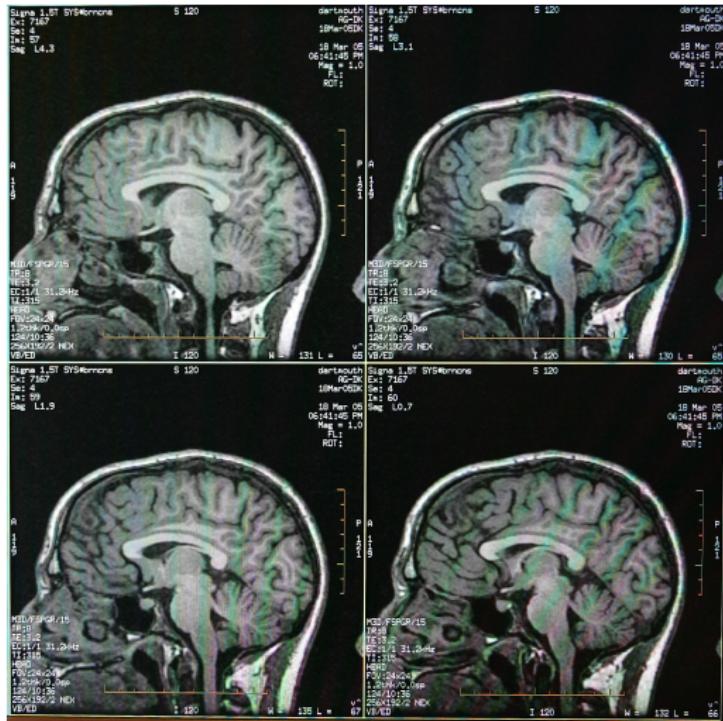
Document collections

<p>SCIENCE: A WEEKLY RECORD OF SCIENTIFIC PROGRESS. ALPHABETIZED INDEX AND TO BIBLIOGRAPHY INDEX OF 1900 VOLUME. PUBLISHED AT THE BREWSTER, NEW YORK.</p> 	<p>SCIENCE A WEEKLY JOURNAL DEVOTED TO THE ADVANCEMENT OF SCIENCE</p> 	<p>SCIENCE A WEEKLY JOURNAL DEVOTED TO THE ADVANCEMENT OF SCIENCE</p> 	<p>SCIENCE A WEEKLY JOURNAL DEVOTED TO THE ADVANCEMENT OF SCIENCE</p> 
<p>SCIENCE A WEEKLY RECORD OF SCIENTIFIC PROGRESS. ALPHABETIZED INDEX AND TO BIBLIOGRAPHY INDEX OF 1910 VOLUME. PUBLISHED AT THE BREWSTER, NEW YORK.</p> 	<p>SCIENCE A WEEKLY JOURNAL DEVOTED TO THE ADVANCEMENT OF SCIENCE</p> 	<p>SCIENCE A WEEKLY JOURNAL DEVOTED TO THE ADVANCEMENT OF SCIENCE</p> 	<p>SCIENCE A WEEKLY JOURNAL DEVOTED TO THE ADVANCEMENT OF SCIENCE</p> 
<p>SCIENCE A WEEKLY RECORD OF SCIENTIFIC PROGRESS. ALPHABETIZED INDEX AND TO BIBLIOGRAPHY INDEX OF 1920 VOLUME. PUBLISHED AT THE BREWSTER, NEW YORK.</p> 	<p>SCIENCE A WEEKLY JOURNAL DEVOTED TO THE ADVANCEMENT OF SCIENCE</p> 	<p>SCIENCE A WEEKLY JOURNAL DEVOTED TO THE ADVANCEMENT OF SCIENCE</p> 	

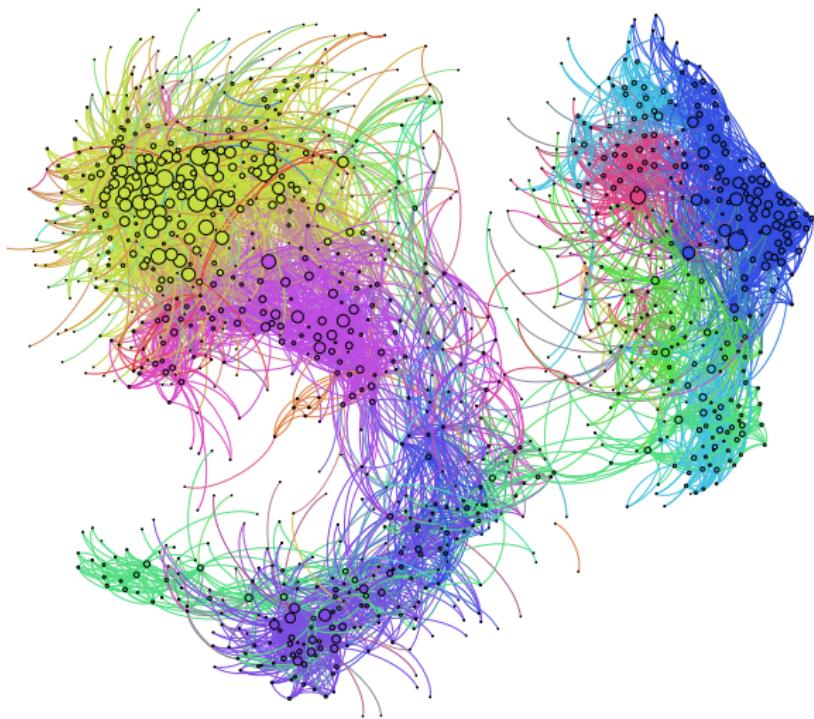
Genomics



Neuroscience



Social networks

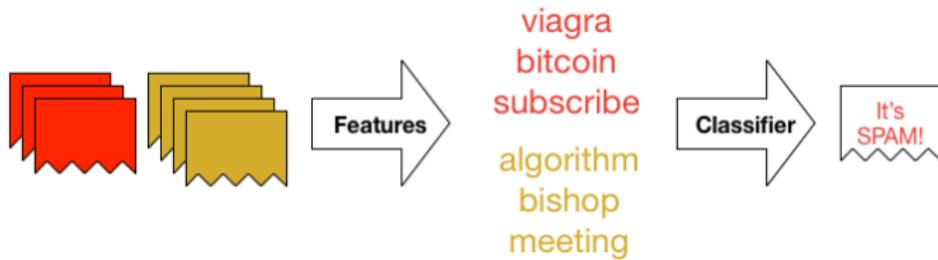


Finance

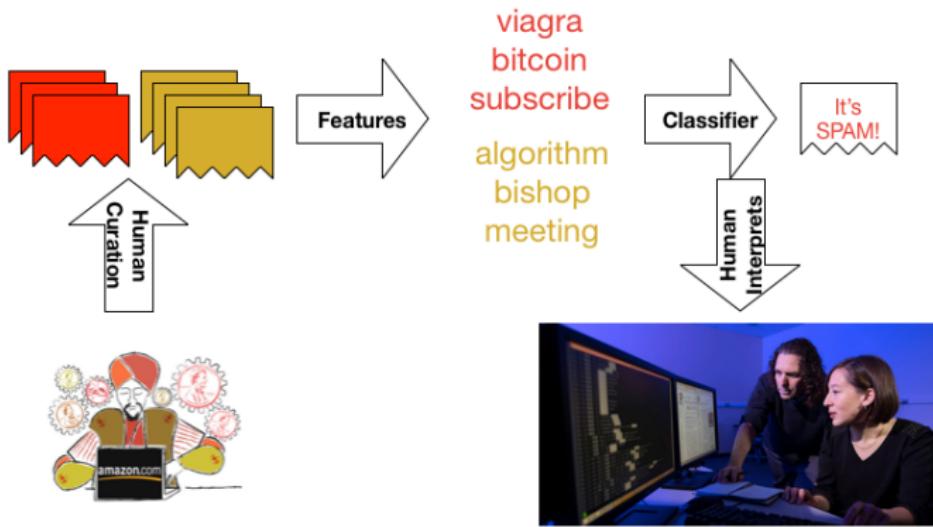


Data are input to algorithms

We can recreate human decisions



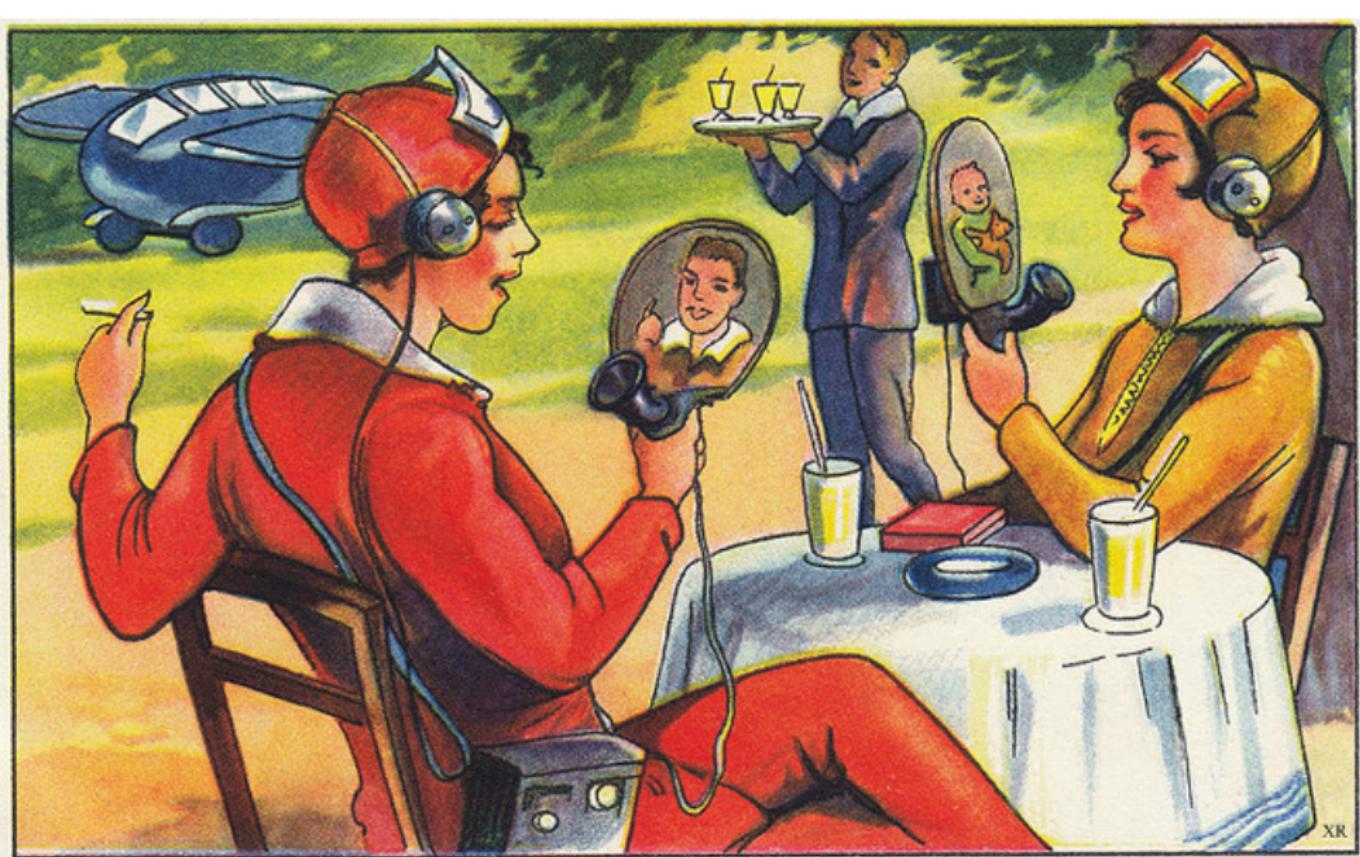
We can recreate human decisions



Algorithms can solve tasks



Waymo driverless car on
the road in Silicon Valley



Predictions: Technology changes but society doesn't



Are we prepared for social changes?

Discussions of AI Need Realism

- You can't solve a task without good data
- Good data often requires humans

Discussions of AI Need Realism

- You can't solve a task without good data
- Good data often requires humans
- Unless task is super-straightforward, also need ML expertise

Discussions of AI Need Realism

- You can't solve a task without good data
- Good data often requires humans
- Unless task is super-straightforward, also need ML expertise
- Effective solutions often require infrastructure

Discussions of AI Need Realism

- You can't solve a task without good data
- Good data often requires humans
- Unless task is super-straightforward, also need ML expertise
- Effective solutions often require infrastructure
 - Explanations
 - Retraining
 - Socio-technical interfaces

Discussions of AI Need Realism

- You can't solve a task without good data
- Good data often requires humans
- Unless task is super-straightforward, also need ML expertise
- Effective solutions often require infrastructure
 - Explanations
 - Retraining
 - Socio-technical interfaces
- DS/ML/AI is **hard** (but rewarding!)

