

# Learning

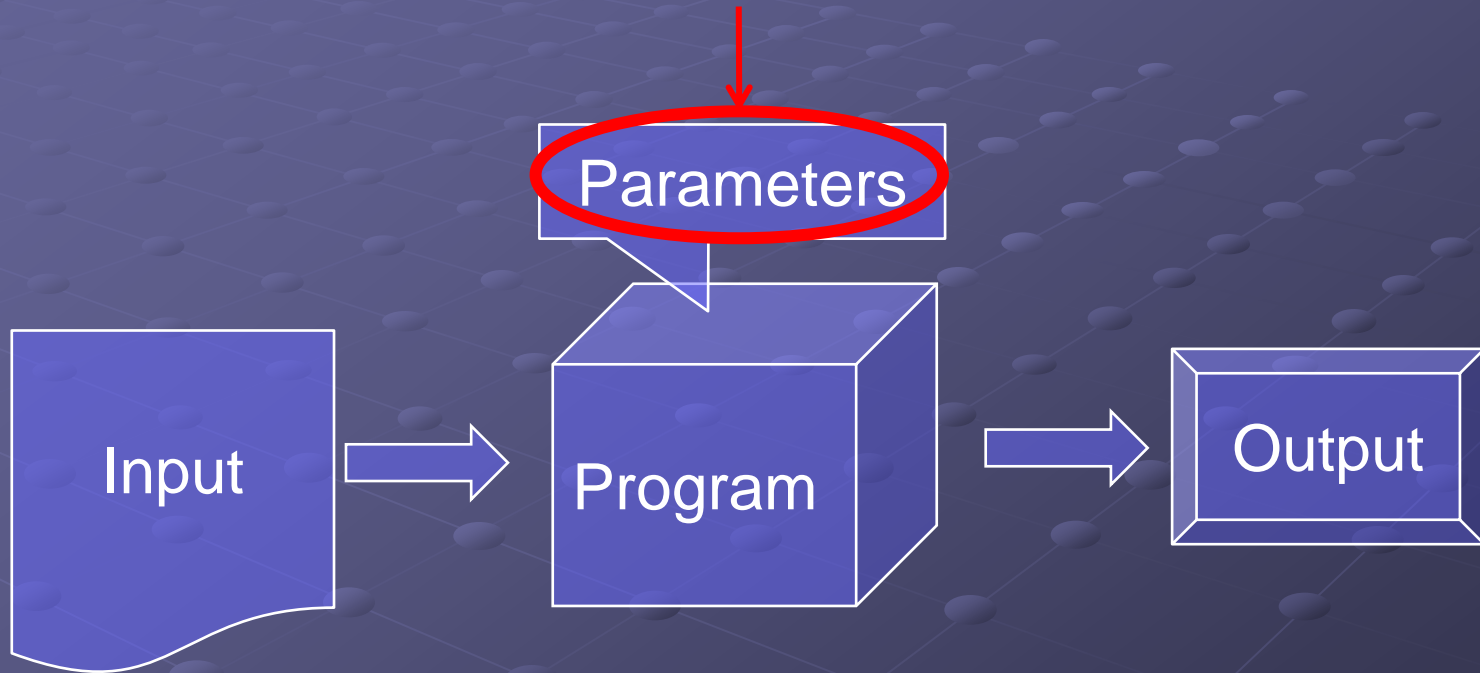
to improve our lives

Daphne Koller  
Stanford University



# Computers Can Learn?

Learned to get desired input/output mapping



- Computers can learn to *predict*
- Computers can learn to *act*

# Many, many, many applications

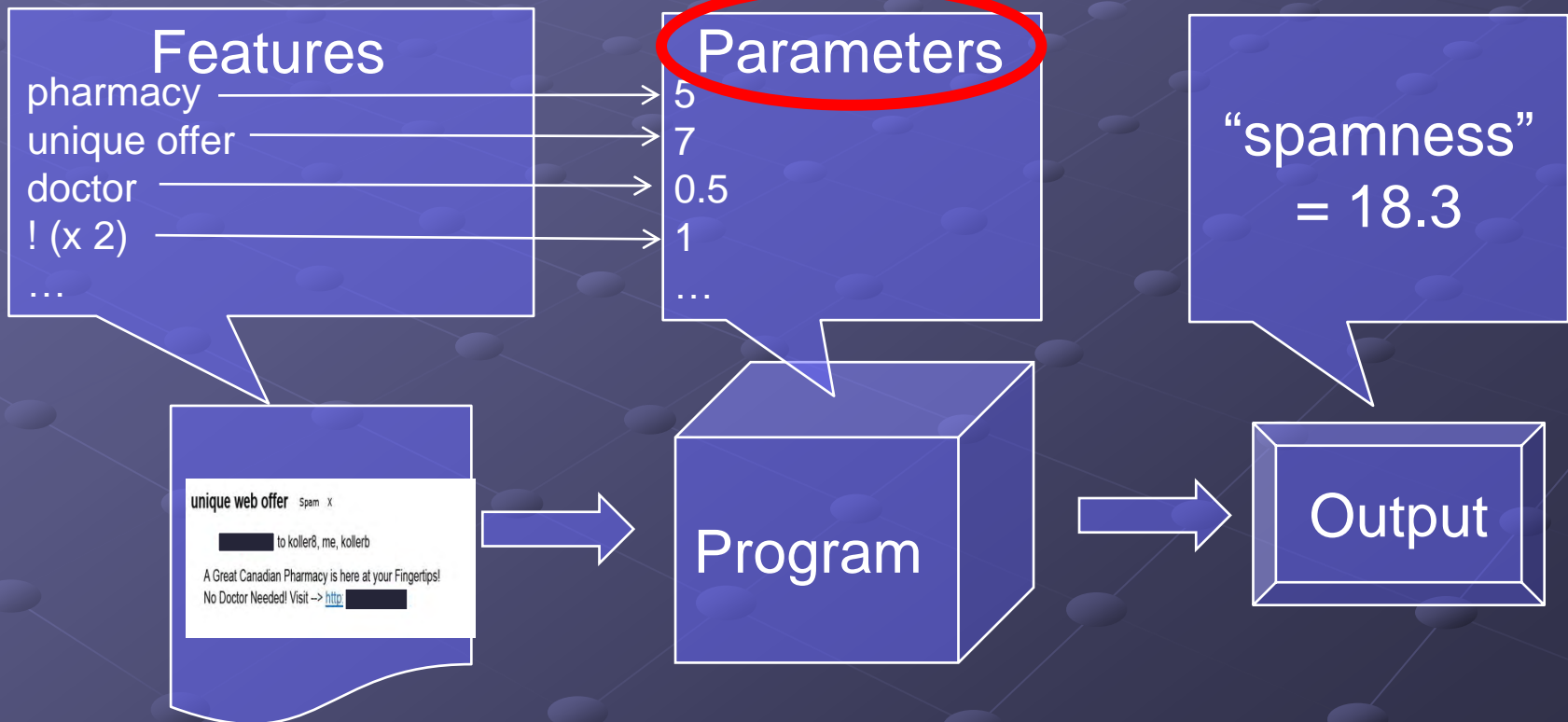
- Speech recognition
- Fraud detection
- Intrusion detection into computer systems
- Image search
- Activity recognition in video surveillance
- Autonomous driving (DARPA Grand Challenge)
- Early epidemic detection
- Cancer subtype classification
- Uncovering basic biological mechanisms
- ....

# Example: Spam Filtering

- Spam email:
  - Comprises 85-95% of email traffic
  - Cost US organizations > \$13 billion in 2007
- Spammers are constantly adapting
  - ⇒ hand-constructed systems bound to fail

# Learning to Detect Spam

- Learned to optimize prediction quality
  - Increase parameters for words appearing in spam email
  - Decrease parameters for words appearing in good email



# Learning to Detect Spam

- Learned to optimize prediction quality
  - Increase parameters for words appearing in spam email
  - Decrease parameters for words appearing in good email
- Can learn in advance
- Online learning adapts to changing trends
- And to personalize to a user's preferences
- Collaborative learning allows learning from other people's data

# Harder: Machine Translation

- Input can't be viewed as a “bag” of words
- Output is not a simple decision (spam / not spam) but a complex sentence
- Machine translation using human-constructed translation rules floundered for decades

The spirit is willing but the flesh is weak.

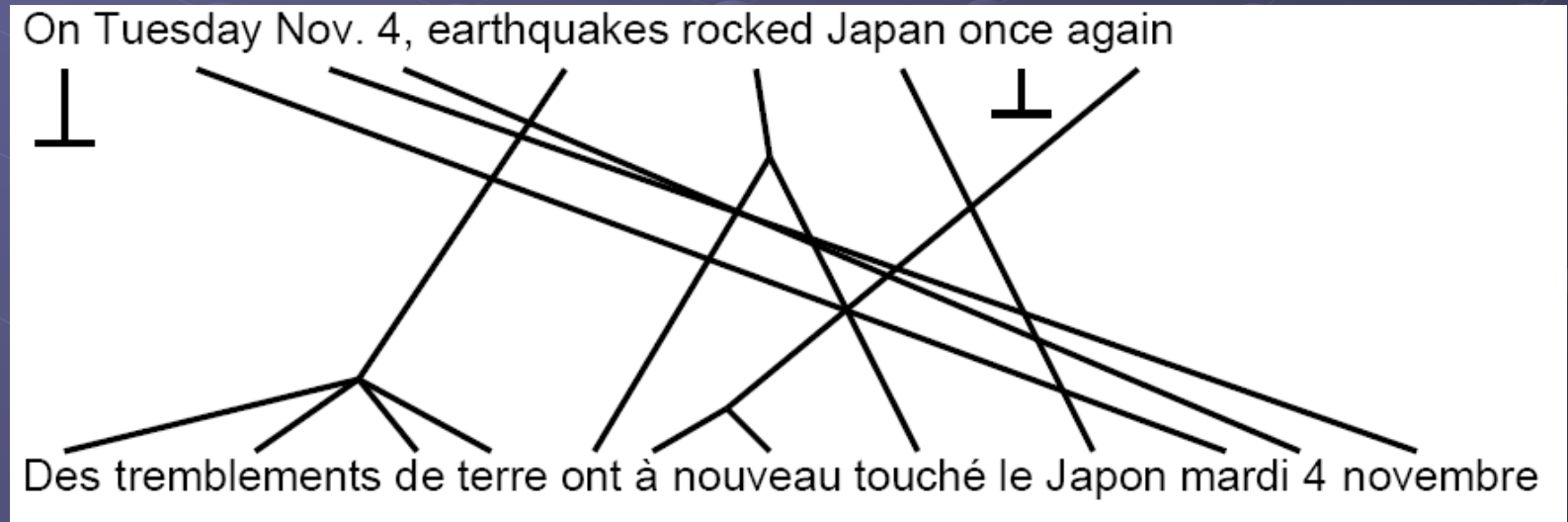


English to Russian and back

The vodka is good but the meat is rotten.

# Harder: Machine Translation

- ML-based machine translation systems
  - Use matched text in two languages to learn matching between words or phrases
  - text in target language to learn what “good” text is like





# Harder: Machine Translation

- ML-based
  - Use match between words in source text and words in target language

**Translate text or webpage**

Enter text or a webpage URL.

Spanish > English swap **Translate**

**Search other languages**

Enter a search phrase in your own language to find information in other languages.  
[Translate and search](#)

**Translation tools**

Get translation with a single click and make your webpage instantly available in other languages.  
[Get translation tools now](#)

**Languages available for translation**

|           |            |
|-----------|------------|
| Albanian  | Indonesian |
| Arabic    | Italian    |
| Bulgarian | Japanese   |
| Catalan   | Korean     |
| Chinese   | Latvian    |
| Croatian  | Lithuanian |
| Czech     | Maltese    |
| Danish    | Norwegian  |
| Dutch     | Polish     |
| English   | Portuguese |
| Estonian  | Romanian   |
| Filipino  | Russian    |
| Finnish   | Serbian    |
| French    | Slovak     |
| Galician  | Slovenian  |
| German    | Spanish    |
| Greek     | Swedish    |
| Hebrew    | Thai       |
| Hindi     | Turkish    |
| Hungarian | Ukrainian  |
|           | Vietnamese |

**Google Translate BETA**

Google Home - [About Google Translate](#)

©2009 Google

On Tuesday No



Des trembleme

” text is like

ardi 4 novembre

# Perception

- Impossible using hand-coded rules
- Example: Automated handwriting recognition
  - Deployed at all 250+ Postal Distribution Centers
  - 25 billion+ letters processed annually
  - > 92% automated processing
  - Hundreds of millions of \$ saved each year



Typical Mail-piece Address Block :

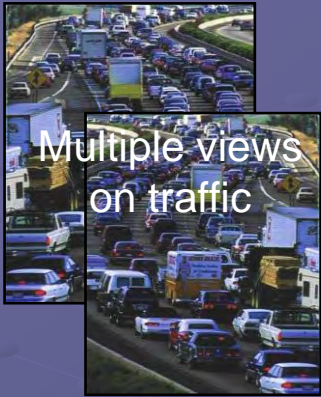
Mrs Knowles  
74 Livingston St  
Buffalo, NY 14213

ZIP and Primary Number Recognition

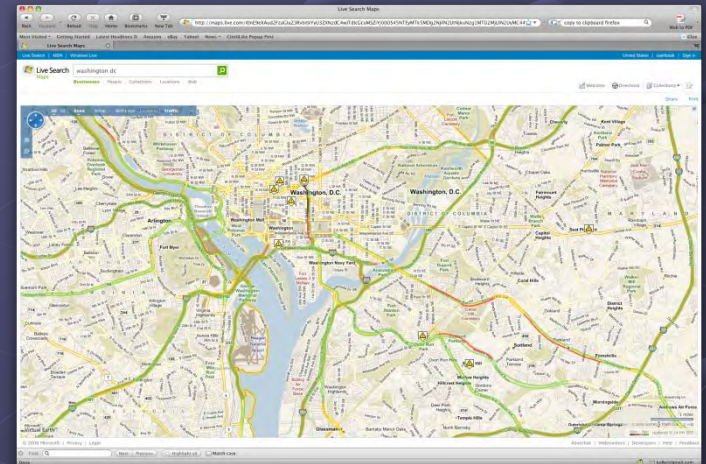
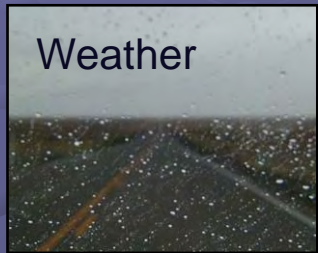
ZIP Code: 14213 Database Query  
Primary Number: 74

| Lexicon Entry (Street Name) | Addon |
|-----------------------------|-------|
| BRADLEY ST                  | 1024  |
| COLONIAL CIR                | 1467  |
| DEWITT ST                   | 1534  |
| GRANT ST                    | 1905  |
| HAMPSHIRE ST                | 2012  |
| HAWLEY ST                   | 1053  |
| LAFAYETTE AVE               | 1347  |
| LIVINGSTON ST               | 1653  |
| MANCHESTER PL               | 1251  |
| PO BOX                      | 0074  |
| POOLEY PL                   | 1128  |
| POTOMAC AVE                 | 1146  |
| PUTNAM ST                   | 1650  |

# Multi-Sensor Integration: Traffic



- Trained on historical data
- Learn to predict current & **future** road speed, including on **unmeasured** roads
- Dynamic route optimization

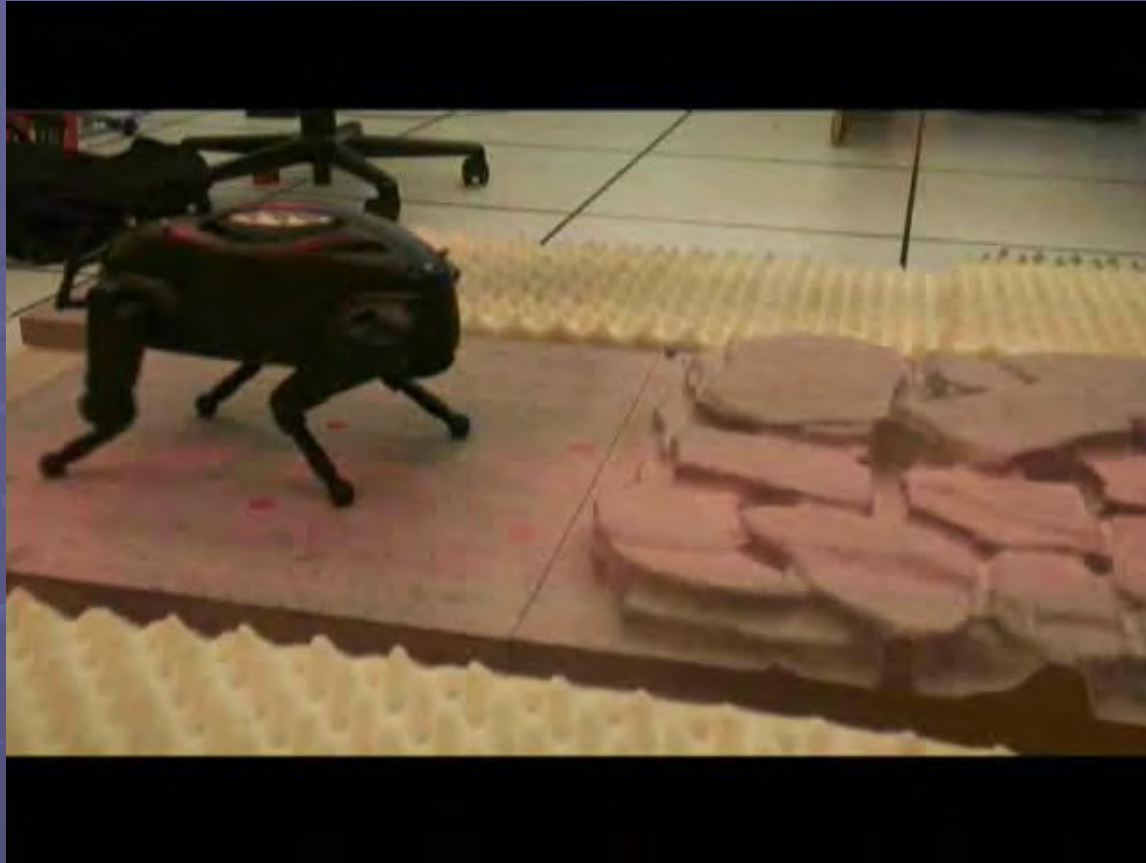


- 195 corridor experiment: accurate to  $\pm 5$  MPH in 85% of cases
- Fielded in 72 cities

Thanks to: Eric Horvitz



# Controlling Complex Systems



This video is available at: <http://www.cra.org/ccc/docs/locslides/quadruped.wmv>

Thanks to: Andrew Ng

# Controlling Complex Systems

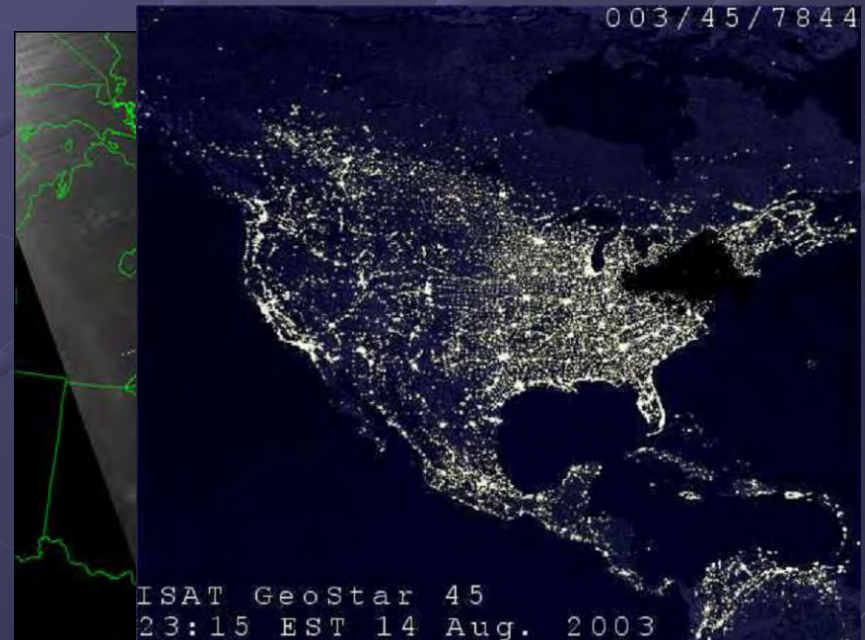


Video available at: <http://www.cra.org/ccs/docs/locslides/helicopter.wmv>

- Learning by emulating a human (apprenticeship)
- ... and by adapting to experience
  - Adjust parameters to reward good behavior

# Future: Smart Power Grid

- Key problem: Get (clean) energy from where it's produced to where it's needed on limited grid
- **Solution: Learning**
  - Perception: predicting current and future demands
  - Control: Make robust and efficient routing decisions



# Medical Diagnosis

- Improve quality of diagnosis:
  - Computer diagnosis systems outperform most doctors
- Allow triage by less-experienced people

The screenshot shows a web browser window with the title 'Applet started'. The browser's address bar and navigation buttons are visible. The website header includes the 'OnParenting' logo (dated May 14 - May 20, 1997), the 'Fidelity Investments' logo (Fidelity Distributors Corporation), and a banner for 'Our home on the web [is where]'. A navigation menu contains links for 'cover', 'contents', 'news', 'experts', 'fun', 'handbook', 'talk', 'find', 'help', and 'feedback'. The main content area is titled 'Describe the child' and includes a text box for instructions: 'There are two ways to search for specific information in OnParenting. In Find by Word, type the word(s) you want to find and get a list of titles relevant to that word. Find by Symptom will help you get information about children's symptoms. Help has tips to target your search.' Below this are two buttons: 'Find by Word' and 'Find by Symptom'. The 'Describe the child' section contains a text box with the instruction: 'Describe the child in the drop-down boxes at the right. Relevant information will appear below.' To the right of this text box are three dropdown menus: 'Age: Toddler', 'Sex: Female', and 'Complaint: Abdominal pain'. Below these are radio button options for 'Localized pain: Can the child localize, or point to, the site of the pain?' with choices: 'No, unable to localize', 'Below the navel to the child's left', 'Above the child's navel', 'Either of the child's sides', 'Below the navel to the child's right', 'Above the navel to the child's right', 'Above the navel to the child's left', and 'Don't Know'. To the right of these options is a 'Results so far' section with a table showing 'Disorder' and 'Relevance' for 'Viral gastroenteritis', 'Psychosomatic pain', 'Urinary tract infection', and 'Other'. At the bottom of the form are four buttons: 'Start Over', 'Review', 'Next>>', and 'Finish'.

Applet started

ON STAGE ESSENTIALS COMMUNICATE FIND

OnParenting May 14 - May 20, 1997

Fidelity Investments Fidelity Distributors Corporation

Our home on the web [is where] click here

cover contents news experts fun handbook talk find help feedback

There are two ways to search for specific information in OnParenting. In Find by Word, type the word(s) you want to find and get a list of titles relevant to that word. Find by Symptom will help you get information about children's symptoms. Help has tips to target your search.

Find by Word

Find by Symptom

**Describe the child** in the drop-down boxes at the right. Relevant information will appear below.

Age: Toddler Sex: Female

Complaint: Abdominal pain

Localized pain: Can the child localize, or point to, the site of the pain?

No, unable to localize

Below the navel to the child's left

Above the child's navel

Either of the child's sides

Below the navel to the child's right

Above the navel to the child's right

Above the navel to the child's left

Don't Know

**Results so far**

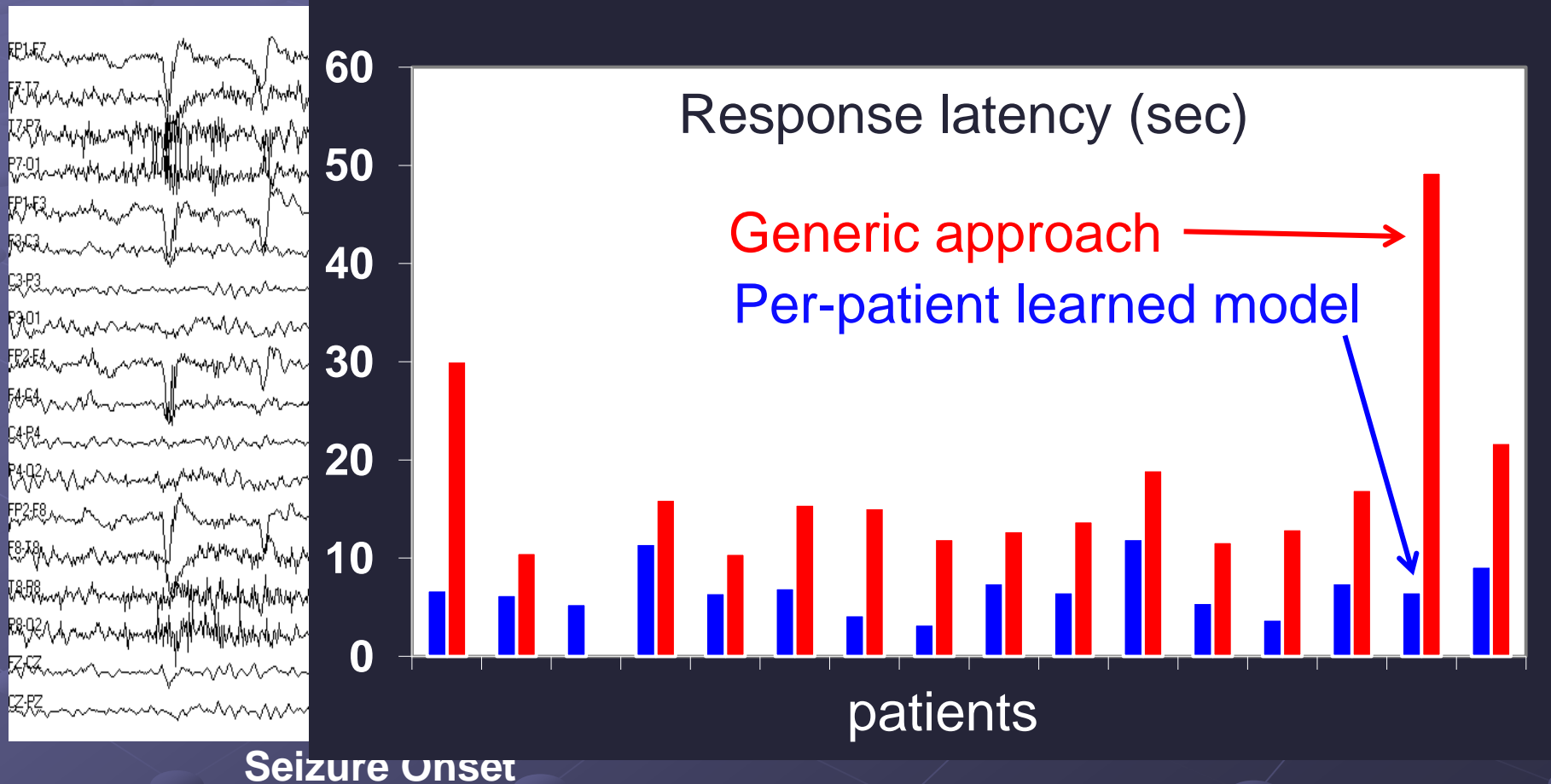
| Disorder                | Relevance            |
|-------------------------|----------------------|
| Viral gastroenteritis   | <input type="text"/> |
| Psychosomatic pain      | <input type="text"/> |
| Urinary tract infection | <input type="text"/> |
| Other                   | <input type="text"/> |

Start Over Review

Next>> Finish

# Medical Intervention

- Patient-specific automatic detection of epilepsy seizures from EEG for real-time intervention



Seizure Onset



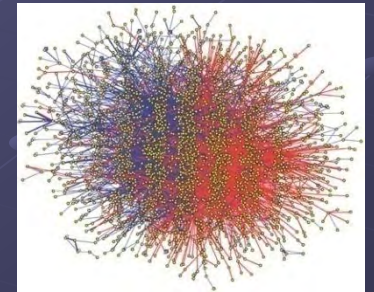
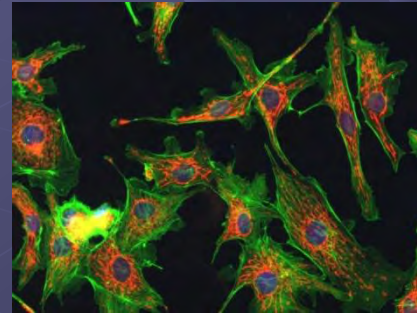
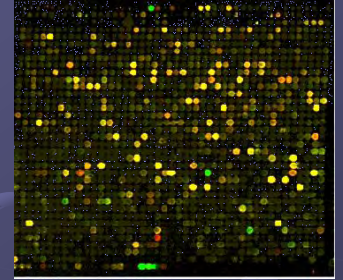
# Medical Intervention

- Patient-specific automatic detection of epilepsy seizures from EEG for real-time intervention
- Reduce frequency of medical errors
  - Learn “standard of care” and detect anomalies
  - Reduce enormous cost: financial and human life
- Home-based systems for tracking of chronic patients for early prediction of complications
  - Reduce pain, suffering, and cost of hospitalization

# Scientific Discovery

## ● New technologies revolutionize biology

- High-throughput sequencing
- Gene expression
- Protein-protein interactions
- Proteomics
- Cellular microscopy
- .....



## ● But how do these help understand & cure disease?

# Our Genes Determine Who We Are

- Humans differ in 0.1% of their DNA
- These differences determine who we are, what diseases we'll get, and which cures will work for us
- Which differences matter?

Diabetes patients



```
...ACTCGGTAGGCATAAATTCGGCCCGGTCAGATTCCATACAGTTTGTACCATGG...  
...ACTCGGTGGGCATAAATTCGGCCCGGTCAGATTCCATACAGTTTGTTCATGG...  
...ACTCGGTAGGCATAAATTCGGCCCGGTCAGATTCCATACAGTTTGTACCATGG...  
:  
...ACTCGGTGGGCATAAATTCGGCCCGGTCAGATTCCATCCAGTTTGTACCATGG...  
...ACTCGGTGGGCATAAATTCGGCCCGGTCAGATTCCATACAGTTTGTTCATGG...
```

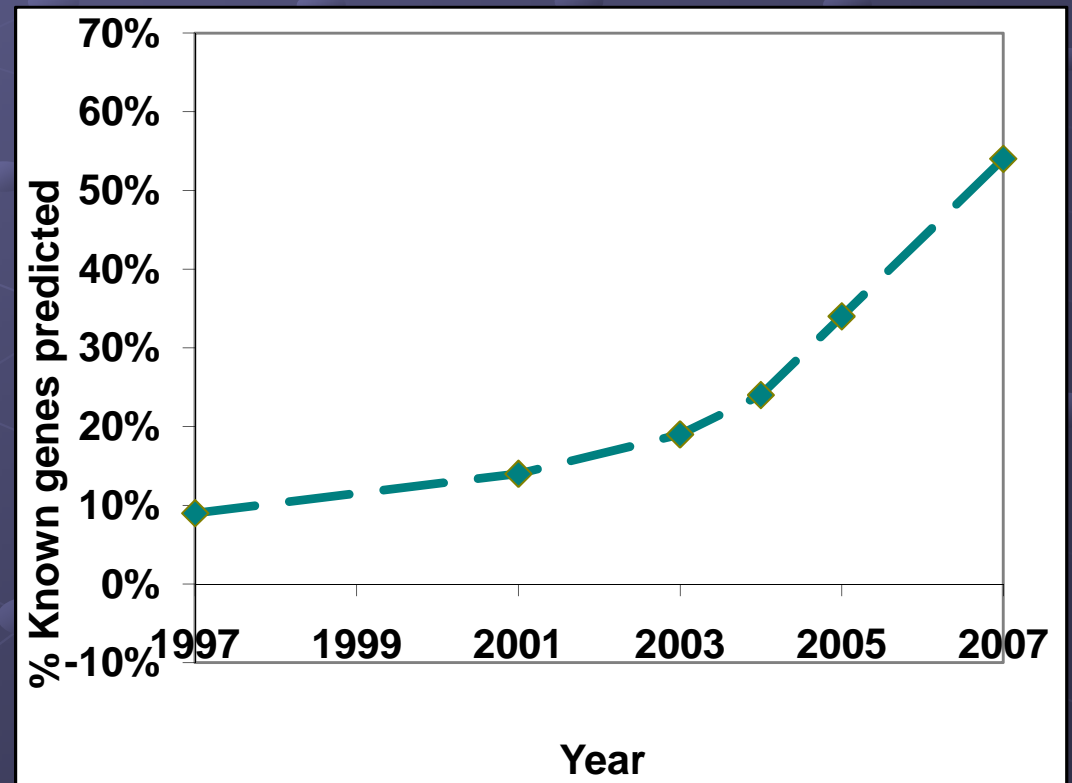
Healthy individuals



```
...ACTCGGTGGGCATAAATTCGGCCCGGTCAGATTCCATCCAGTTTGTTCATGG...  
...ACTCGGTGGGCATAAATTCGGCCCGGTCAGATTCCATCCAGTTTGTACCATGG...  
...ACTCGGTGGGCATAAATTCGGCCCGGTCAGATTCCATCCAGTTTGTACCATGG...  
:  
...ACTCGGTGGGCATAAATTCGGCCCGGTCAGATTCCATCCAGTTTGTACCATGG...  
...ACTCGGTGGGCATAAATTCGGCCCGGTCAGATTCCATCCAGTTTGTTCATGG...
```

# Where Are the Genes?

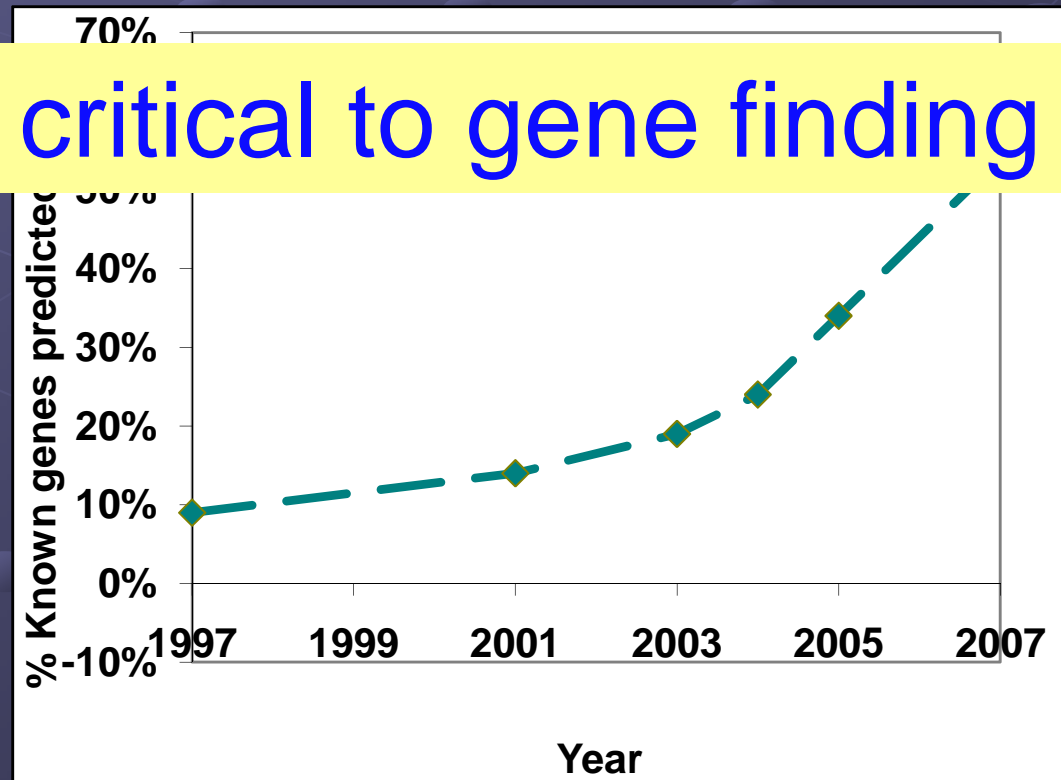
- Only 5% of DNA appears to play functional role
- To understand which genetic changes matter, we need to find the functional pieces, such as genes
- Train model using known genes
- Learn what DNA sequences characterize them



# Where Are the Genes?

- Only 5% of DNA appears to play functional role
- To understand which genetic changes matter, we need to find the functional pieces, such as genes
- Train model using known genes
- Learn what DNA sequences characterize them

Machine learning critical to gene finding



# Future: Smart Healthcare

- Evidence-based medicine: Learn what works
- ... at personalized level: What works **for me**
- **Learn** mapping from individual genotype and other factors to disease risk and drug suitability

# Machine Learning = Computing on Steroids

Challenging  
Application

Machine Learning

Data

# Machine Learning = Computing on Steroids

- ML core technology for prediction and decision
- Makes possible applications where other methods simply don't work
  - Perception
  - Personalization
  - Dynamic adaptation
- Can improve almost any application
- A little bit of learning goes a long way