What’s something that excites you about CS?
a topic, algorithm, language, application…
What is computer science research?
STANFORD, Calif. — You may think you can find almost anything on the Internet.

But even as images and video rapidly come to dominate the Web, search engines can ordinarily find a given image only if the text entered by a searcher matches the text with which it was labeled. And the labels can be unreliable, "fuzzy" instead of "rabbit" or simply nonexistent.

To eliminate those limits, scientists will need to create a new generation of visual search technologies — or else, as the Stanford computer scientist Fei-Fei Li recently put it, the Web will be in danger of "going dark."

Now, along with computer scientists from Princeton, Dr. Li, 36, has built the world's largest visual database in an effort to mimic the human vision system. With more than 14 million labeled objects, from obsidian to orangutans to ocelots, the database has become a vital resource for computer vision researchers.

Stanford Researcher Finds Lots of Leaky Web Sites

BY SOMINI SENGUPTA
October 11, 2011 6:32 PM

The Web is porous. Remarkable information trickles in from everywhere. It also sometimes spills out without its users knowing exactly where or how.

Take for instance these findings, released on Tuesday by computer scientists at Stanford University. If you type a wrong password into the Web site of The Wall Street Journal, it turns out that your e-mail address quietly slips out to seven unrelated Web sites. Sign on to NBC and, likewise, seven other companies can capture your e-mail address. Click on an ad on HomeDepot.com and your first name and user ID are instantly revealed to 13 other companies.

These findings, released by the Center for Internet and Society at Stanford Law School, are among the leaks found on 185 top Web sites. They serve to buttress what privacy advocates have long
What are some of the first things that come to mind when you think of CS research?
What will this course achieve?
Your experience in CS 197

Work on bleeding-edge topics now, rather than in two years

Fashion a project that you can publish as a work-in-progress or workshop paper

Find an onramp to research in the department, and to research and advanced development in industry
Alex Tamkin

2014-2018       BS, Stanford (CS)
2018-now        PhD, Stanford (CS)

Fan of ceramics, photography, cooking, puns
(call me Alex)
My Research History

First research project:
CURIS at Stanford after freshman year! HCI Group
My Research History

Junior Year

NLP Research—“natural language inference”

<table>
<thead>
<tr>
<th>Premise</th>
<th>Relation</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>the order compelled him to appear as a witness</td>
<td>entails</td>
<td>he appeared as a witness</td>
</tr>
<tr>
<td>we have missed an opportunity to examine the art</td>
<td>contradicts</td>
<td>we have examined the art market today</td>
</tr>
<tr>
<td>market today</td>
<td>permits</td>
<td>Mr Odinga had changed his plans</td>
</tr>
<tr>
<td>Mr Odinga had not been forced to change his plans</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Examples from SCI randomly chosen from the validation set. Each row contains a triplet formed by a premise (left column), a hypothesis (right column), and a label specifying one of the three possible relations (entails, contradicts, permits) holding between premise and hypothesis. The last row contains an example of a probabilistic implicative (see the main text).
My Research History

After Senior Year: Google Brain, ML for detecting exoplanets
# My Research History

**PhD: Unsupervised Machine Learning**

<table>
<thead>
<tr>
<th>Model Architecture</th>
<th>Pretraining Objective</th>
<th>Transfer Method</th>
<th>Downstream Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Images" /></td>
<td><img src="image2.png" alt="Images" /></td>
<td><img src="image3.png" alt="Images" /></td>
<td><img src="image4.png" alt="Images" /></td>
</tr>
</tbody>
</table>

- **Unlabeled Data**
- **NATURAL IMAGES**
- **SPEECH**
- **TEXT**
- **MEDICAL IMAGING**
- **SENSORS**
- **CAPTIONED IMAGES**
What about you?

What excites you about CS research
(topic? person? idea?)
Today

What is research, vs. industry?
Research mindset
How does this course work?
Computer science research

What is the goal of research?
Why has it driven major innovations in computing?
What separates research from advanced development?
A Tale of Three Turing Awards
Hennessy and Patterson: RISC

Computer architecture was increasing in complexity, in order to enable more and more advanced computation.

Everyone thought that increasingly powerful processors needed increasingly complicated instruction sets to take advantage of them.
Hennessy and Patterson: RISC

“No, let’s do it this way instead:” have a very simple instruction set. That way you can compare performance, optimize, and prevent errors.

This became known as Reduced Instruction Set Computer (RISC). It led to a sea change in architectures, and the founding of multiple major silicon valley companies.
Engelbart: interactive computing

When computers originated, they were used for, well, computing: calculating mathematical functions.

This meant that computers were seen as most appropriate for slow, batch interaction, shared by entire teams.
“No, let’s do it this way instead:” computing should be used as a tool for thought. We must move from batch-style computing to interactive computing.

His result was the “Mother of All Demos”: mouse, hypertext, bitmapped screens, collaborative software, and more.

This led to Xerox Star. Steve Jobs saw it, was wow’ed, and infused the ideas into the Mac.
LeCun, Hinton, Bengio: deep learning

The idea of neural networks had been around for fifty years, but unsuccessful. Major AI figures had trashed it, even proving that early versions had very limited expressiveness.

Instead, machine learning was based on other models, for example the support vector machine and graphical models. Neural networks did not perform well.
“No, let’s do it this way instead:” these networks learn extremely complex functions, so they need much more data than existing machine learning approaches, GPUs to train, and algorithms to enable them to learn more effectively.

Around 2010, these models began smashing records in speech and image recognition. They are now dominant in ML.
Not all research wins Turing Awards. But…

It all follows this same formula —

An implicit assumption: Industry and other researchers all thought one way about a problem

“No, let’s do it this way instead:” The researcher offered a new perspective that nobody had ever considered or made feasible before. They proved out their idea as the better approach.
And now, a definition.

Research introduces a fundamental **new idea** into the world.

Examples:

Simple instruction sets for complex computer architecture
Computing that is interactive, not batch
Algorithms needed to make deep learning effective

These ideas did not exist in any mature or well-articulated way before their creators developed them.

If the idea is already in the world, for example published by someone else, it is not considered **novel**, and thus not research.
Before: small computer vision datasets
After: HUGE computer vision dataset, and algorithms to utilize it

Before: programmers manually reserve resources for cloud computing
After: programmers provide needs, software allocates resources

Before: programmers manually reserve resources for cloud computing
After: programmers provide needs, software allocates resources

Before: underwater robots should look and feel like boats
After: humanoid underwater robotics
Research creates industry

Google
- PageRank algorithm

Sun Microsystems
- Stanford University Network workstation (SUNet)

sgi
- Computer graphics architectures

Coursera
- Online education

VMware
- Computer virtualization
Industry and research
Industry vs. research

What makes other start-ups and industry different than research?
If the core idea already exists, but needs to be refined in order to see success…it might be important, but it’s not research.
Industry vs. research

Companies can and do engage in development that is research…

MapReduce and Spanner at Google
Kinect at Microsoft

…but typically companies are working to scale out ideas that exist.

Landay, 2000s: activity sensing
Credit because he developed the concept and popularized it

Apple, 2010s: Apple Watch
Credit because they engineered it to work and launched it
Implication: by doing research, you are living about 15 years in the future.
(An incomplete list of) research areas in computer science
Flavors of CS research

Computer science is a field held together by a shared phenomenon of interest: computing.

This sets it apart from some other fields, which are drawn together by a shared theory or shared methodology. While this is a simplification, it is a helpful first cut:

- Psychology: methodology of human subjects experiments
- Math: methodology of formal proof
- Anthropology: methodology of participant observation
- Sociology: shared theories — functionalist perspective, conflict perspective, symbolic interactionist perspective
Architecture
Artificial intelligence
Computational biology
Computer graphics
Computer security
Computer systems
Computer vision
Data science
Education

Human-computer interaction
Machine learning
Natural language processing
Networking
Operating/distributed systems
Programming systems/verification
Robotics
Theory
Topic: artificial intelligence

Architecture

Artificial intelligence

Computational biology

Computer graphics

Computer security

Computer systems

Computer vision

Data science

Education

Human-computer interaction

Machine learning

Natural language processing

Networking

Operating/distributed systems

Programming systems/verification

Robotics

Theory
Topic: computer systems

Architecture
Artificial intelligence
Computational biology
Computer graphics

Computer security

Computer systems
Computer vision
Data science
Education

Human-computer interaction
Machine learning
Natural language processing

Networking

Operating/distributed systems
 Programming systems/verification
Robotics
Theory
Topic: theory

Architecture
Artificial intelligence
Computational biology
Computer graphics

**Computer security**

Computer systems
Computer vision
Data science
Education

Human-computer interaction

**Machine learning**

Natural language processing
Networking

Operating/distributed systems

**Programming systems/verification**

Robotics

Theory
Method: engineering

**Architecture**
- Artificial intelligence
- Computational biology

**Computer graphics**

**Computer security**

**Computer systems**
- Computer vision

**Data science**

**Education**

**Human-computer interaction**
- Machine learning
- Natural language processing

**Networking**

**Operating/distributed systems**

**Programming systems/verification**

**Robotics**

**Theory**
<table>
<thead>
<tr>
<th>Method: probability and modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture</td>
</tr>
<tr>
<td>Artificial intelligence</td>
</tr>
<tr>
<td>Computational biology</td>
</tr>
<tr>
<td>Computer graphics</td>
</tr>
<tr>
<td>Computer security</td>
</tr>
<tr>
<td>Computer systems</td>
</tr>
<tr>
<td>Computer vision</td>
</tr>
<tr>
<td>Data science</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Human-computer interaction</td>
</tr>
<tr>
<td>Machine learning</td>
</tr>
<tr>
<td>Natural language processing</td>
</tr>
<tr>
<td>Networking</td>
</tr>
<tr>
<td>Operating/distributed systems</td>
</tr>
<tr>
<td>Programming systems/verification</td>
</tr>
<tr>
<td>Robotics</td>
</tr>
<tr>
<td>Theory</td>
</tr>
</tbody>
</table>
Method: formal reasoning and proof

Architecture

Artificial intelligence

Computational biology

Computer graphics

Computer security

Computer systems

Computer vision

Data science

Education

Human-computer interaction

Machine learning

Natural language processing

Networking

Operating/distributed systems

Programming systems/verification

Robotics

Theory
Method: design

Architecture
Artificial intelligence
Computational biology
Computer graphics
Computer security
Computer systems
Computer vision
Data science
Education

Human-computer interaction
Machine learning
Natural language processing
Networking
Operating/distributed systems
Programming systems/verification
Robotics
Theory
Method: empirical measurement and hypothesis testing

Architecture
Artificial intelligence

**Computational biology**
Computer graphics
Computer security
Computer systems
Computer vision

Data science

Education

**Human-computer interaction**
Machine learning
Natural language processing
Networking
Operating/distributed systems
Programming systems/verification
Robotics
Theory
Research mindset
Research is different than your usual coursework.

Coursework tends to be very clearly defined. Research tends to be exploratory and iterative.
“I like” from summer research:

“The free-form structure of our project”

“The freedom to choose the questions and methods I find interesting”

“The independence I got in establishing a research direction”

“That I have had the opportunity to do a lot of self guided research and reading. I feel very free to shape parts of my learning and research experience.”

“I wish” from summer research:

“That there was more structure or well-defined expectations.”

“I had a clearer idea or more deliverables and felt the barrier of being unfamiliar with certain parts of the project or coming on late less.”

"I had been able to narrow my scope a little earlier"

Research is a new and different skill. Embrace and navigate through the uncertainty.
How this course works
Section 01
Learning goals

Execute a **first research project** at the scale that can be submitted to a workshop or work-in-progress at a top-tier conference.

Understand the **major research topics** currently active in your area. Be able to read a research paper and perform a literature review in that area.

Apply **vectoring and velocity skills** for navigating the open-ended nature of research.

Design and execute an appropriate **evaluation** of your method.

Write a paper and engage in the **peer review** process.
Is this course right for me?

CS 197 is the best fit if you’re…

- Interested in working on bleeding-edge problems before you’re a senior or coterm
- Done with CS 106B and ideally taking/taken CS 107

It’s not the best fit if you’re…

- Looking for a research area that we don’t cover yet
- A senior or coterm with the coursework to enable you access to research opportunities already
Course application

For this third offering of the course, we will have space for twelve students per section: AI, HCI, and CompBio.

https://forms.gle/5kWSX6P1K4ysgPF46

Due at 6:00pm today

Decisions + waitlist announced tonight

Sections start tomorrow
Research project

This class is structured around a quarter-long research project. The project is completed in groups of **3-4 students** within a section. TAs will offer project options tailored to each section and the students’ interests within the section. These projects are designed to be accessible to you, of interest to the research community, and achievable within the timeline of the course.
Groups and projects

Form project teams and align with topics in section during Week 2.

You can pick your groups, and your group can pick a project from at the provided list:

The TAs will be the advisors on the project and help scope your direction and next steps each week.

…but you’ll also have some freedom to evolve the shape of the project.
Assignments

Assignments offer waypoints in support of the project.

Assignment 1 (individual): learning about the project area, and learning how to read a paper
Assignment 2: literature review
Assignment 3: introduction/proposal draft
Assignment 4: experiments and evaluation
Assignment 5: draft paper and peer review
Sections

We have three sections: HCI, CompBio, and AI. Each section is led by a PhD student who (1) is doing research in that area, and (2) has been selected for their mentorship skills.

HCI (Nava Haghighi): TBD
AI (Alex Tamkin): Thursdays 3:30pm-4:30pm
CompBio (Kelly Cochran): Thursdays 11:00am-12:00pm

TAs, introduce yourselves!

When you are admitted to the class, you are admitted to a particular section. There are twelve spots per section.
What’s after CS 197?

Our goal is for CS 197 to be an onramp for you to research in Computer Science. We will:

Have opportunities for you to continue to work on the project if desired through CS 197A / CS 199 in future quarters, where you continue to meet with your groupmates.

Perform outreach to labs in CS or at Stanford to help introduce you so you can work on research projects after demonstrating excellence here.

Support you in submitting your work to flagship conferences, and connect you with funding opportunities to travel to present the work.
Assignment 1

Read a Paper

Work on Starter Task

Submit project interest form
Course resources

**Course website**: lecture slides, assignment specs, logistics, …

**Slack**: All other communication (feel free to message us!)

**Canvas**: Submit assignments

We’ll send out a Slack invitation after lecture to cs197.slack.com
Questions?
Computer Science Research

Slide content shareable under a Creative Commons Attribution-NonCommercial 4.0 International License.