Language & Technology

Syllabus

Alëna Aksënova & Aniello De Santo

Stony Brook University
alena.aksenova@stonybrook.edu
aniello.desanto@stonybrook.edu

Jan 28, 2019

Organizational Information

Course	Language and Technology
Course#	LIN 120
Room	Humanities 1006
Time - Session 1	MW 10:00–10:53
Time - Session 2	MW 11:00–11:53
Website	Cocalc and Blackboard

Instructor	Alëna Aksënova	Aniello De Santo
Office hours	W 1:00 – 4:00	M 12:00 – 2:00
		F 12:00–1:00
Office	SBS N210	SBS N232

TAs	Pablo Lopes Alonso & Kalina Kostyszyn & Jun Lyu
Undergraduate TAs	Jessica Ju, Zhan Peng Zheng, Cody St. Clair
	Kathryn Chen, Elizabeth Lei, Jack Jiang

See the Blackboard course page for more details and announcements.

Bulletin Description

An introduction to how computers process language and solve language-related tasks. This course discusses the language technologies of our daily life — spam filtering, machine translation, and many more — and shows how they work under the hood. The course explores a variety of issues: Why do computers do well in some areas (spell checking) yet fail miserably in others (essay grading)? Will we ever have perfectly fluent AIs as depicted in science fiction? And how will these technological advances impact the role of language in our society? Students will also acquire basic programming skills and write scripts for simple language tasks. No previous training in mathematics or computer science required.

SBC: TECH

3 credits

An Experiment

1. Open some chat or messaging app on your phone.
2. Don’t type anything.
3. Instead, click the second word suggestion
   (the one in the middle).
4. Keep doing this.
5. Did you get a reasonable sentence of English?

An Experiment

1. Open some chat or messaging app on your phone.
2. Don’t type anything.
3. Instead, click the second word suggestion
   (the one in the middle).
4. Keep doing this.
5. Did you get a reasonable sentence of English?

   I am a beautiful person who is the best of luck to you by the way to get the best of luck to you by the way to get the best of luck to you by the way to get the . . .

The Big Take-Home Message

Current language technology is mostly smoke and mirrors

Questions and Topics

- How do computers process language?
- Why do they succeed in some areas (spell checker, spam filter), yet fail miserably in others (translation, poetry)?
- Will we ever have conversant AIs as depicted in science fiction (2001, Star Trek, Blade Runner, Her, Ex Machina, System Shock)?
- Can computers provide new answers to long-standing questions of linguistics and philology?
- How are language communities affected by these new technologies?

Teaching Goals

- **Basics of Programming and Computer Science**
  - understand the importance of algorithms and data structures
  - conceptualize linguistic problems in computational terms
  - basic programming skills in Python
- **Cognitive Science**
  - familiarity with notions of artificial intelligence
  - understand how and why humans and computers differ in their linguistic abilities
- **Digital Humanities and Social Science**
  - work with text corpora
  - use computational tools for humanities (stylistic analysis, tracking social developments via corpora)
  - understand the role of Big Data in computational linguistics
  - awareness of the dangers of computational linguistics (surveillance, language death)

Teaching Goals

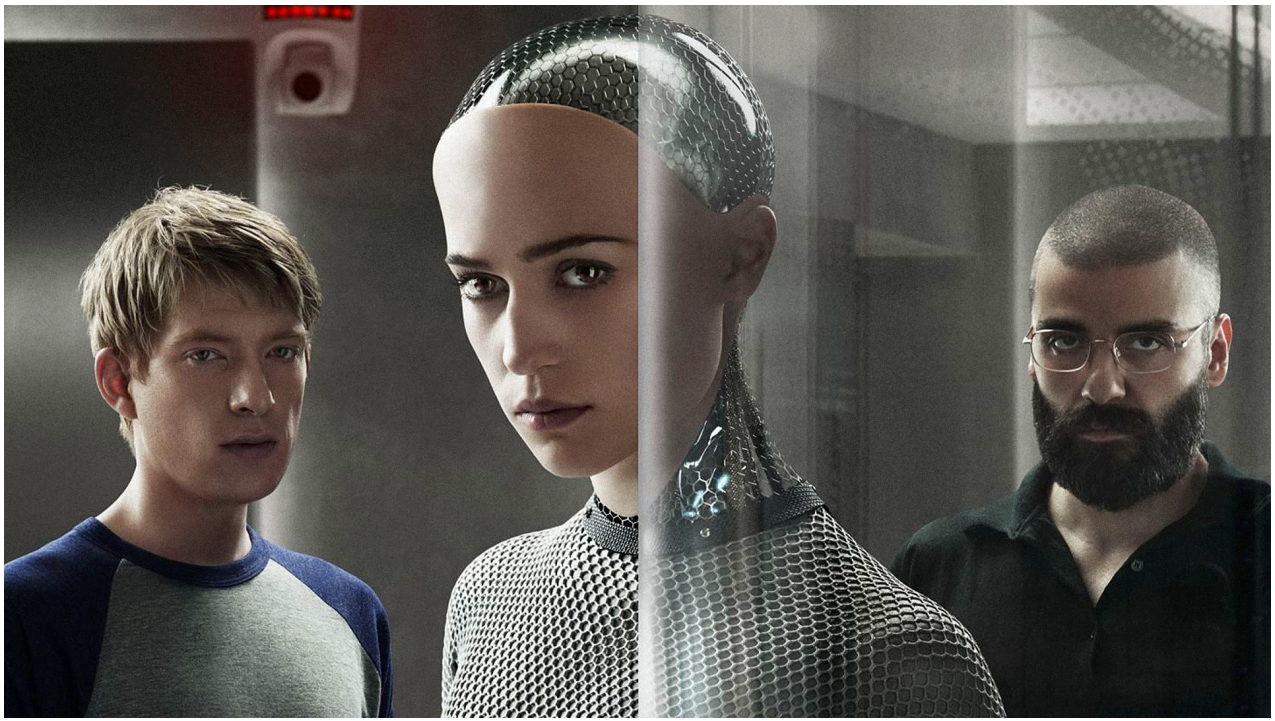
- **Basics of Programming and Computer Science**
  - understand the importance of algorithms and data structures
  - conceptualize linguistic problems in computational terms
  - basic programming skills in Python
- **Cognitive Science**
  - familiarity with notions of artificial intelligence
  - understand how and why humans and computers differ in their linguistic abilities
- **Digital Humanities and Social Science**
  - work with text corpora
  - use computational tools for humanities (stylistic analysis, tracking social developments via corpora)
  - understand the role of Big Data in computational linguistics
  - awareness of the dangers of computational linguistics (surveillance, language death)

Teaching Goals

- **Basics of Programming and Computer Science**
  - understand the importance of algorithms and data structures
  - conceptualize linguistic problems in computational terms
  - basic programming skills in Python
- **Cognitive Science**
  - familiarity with notions of artificial intelligence
  - understand how and why humans and computers differ in their linguistic abilities
- **Digital Humanities and Social Science**
  - work with text corpora
  - use computational tools for humanities (stylistic analysis, tracking social developments via corpora)
  - understand the role of Big Data in computational linguistics
  - awareness of the dangers of computational linguistics (surveillance, language death)

Benchmark

By the end of the course, the following scene from *Ex Machina* should seem rather trivial to you.



Prerequisites

- **What You Need**
  - ability to operate a computer
    (use a web browser, install software, edit text files)
  - willingness to play around with open-ended problems
- **What You WON'T Need**
  - programming experience
  - math (except for addition, multiplication and fractions)
  - linguistics (LIN 101 helps a bit, though)

Three Types of Instruction

Monday	standard lecture on language technology (taught by **Aniello**)
Wednesday	programming sessions in Python (taught by **Alëna**)
Recitation	recap material with your TAs

Session	Mini-quiz?	Laptop?	Attendance?
Monday	yes	not recommended	recommended
Wednesday	no	recommended	recommended
Recitation	no	recommended	mandatory

Echo Video Recordings

- Video recordings of all lectures will be made available online.
- But the system is flaky, don't rely on it.

Three Types of Instruction

Monday	standard lecture on language technology (taught by **Aniello**)
Wednesday	programming sessions in Python (taught by **Alëna**)
Recitation	recap material with your TAs

Session	Mini-quiz?	Laptop?	Attendance?
Monday	yes	not recommended	recommended
Wednesday	no	recommended	recommended
Recitation	no	recommended	mandatory

Echo Video Recordings

- Video recordings of all lectures will be made available online.
- But the system is flaky, don't rely on it.

Grading Components

Class Participation (10%)

- Both in class and **online**!
- **Examples**:
  - ask questions
  - help fellow students
  - link to relevant online materials
  - ⋮
- **Why?**
  - Encourages you to ask questions.
  - Helping others is a great way of learning.
  - We want to have some fun, too.

Grading Components [cont.]

Mini-Quizzes (30%)

- at the beginning of the Monday Lectures
- apply techniques discussed in lecture
- questions template in **online quiz pool** (on CoCalc)
- only pass-fail grading (0 points VS 1 point)
- missed quiz is an automatic fail (but 1 fail per semester is dropped)
- **Why?**
  - We want you to learn skills and techniques, not memorize definitions.
  - Quizzes force you to self-assess how much you are getting out of the class.

Grading Components [cont.]

Python Exercises (30%)

- once per week
- programming in **Python**
- assigned on Wednesday at 11:59pm
- due the following Tuesday at 11:59pm
- assigned and collected through CoCalc
- only pass-fail grading (0 points VS 1 point)
- no late hand-ins (more on that later)
- **Why?**
  - Learning programming is like learning a new language
    $\Rightarrow$ needs constant practice
  - Even a little bit of programming experience is incredibly useful.

Grading Components [cont.]

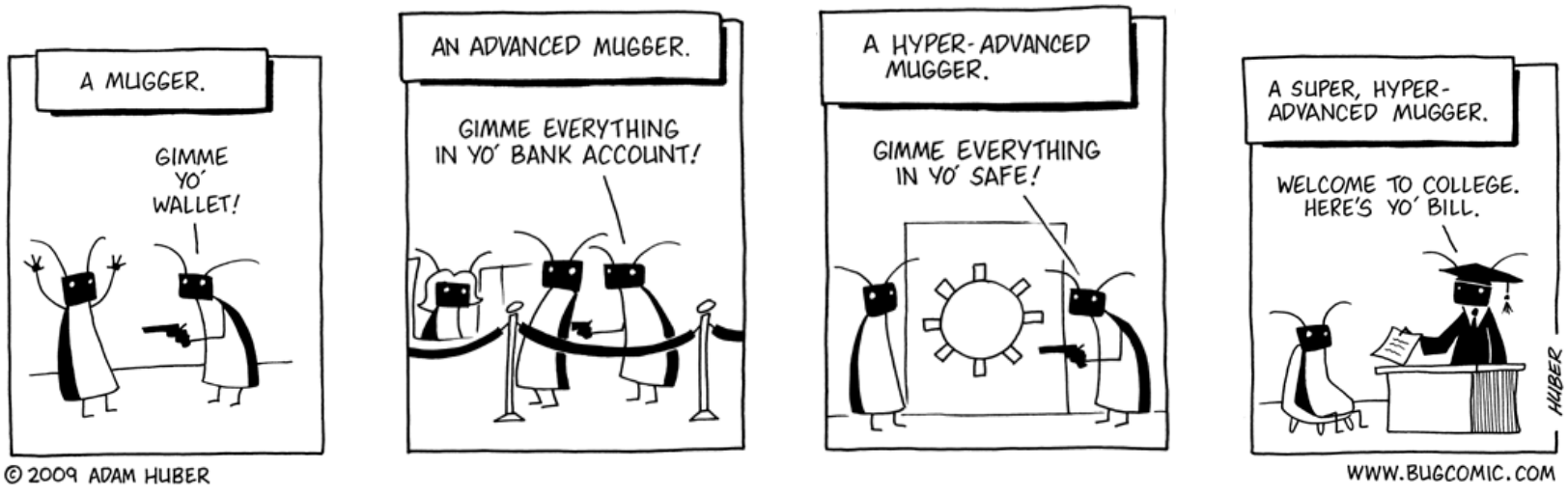
Midterm (30%)

- Tentatively in **Week 7** (during recitation)
- Some theory questions (like the in-class quizzes)
- Some pen-and-paper coding assignment
- **Why?**
  - Force you to check how your study method is working, and eventually correct course
  - Pen-and-Paper coding helps focus on solving the problem and not on the minor details of the coding language (i.e. Python).

Dealing with Fails

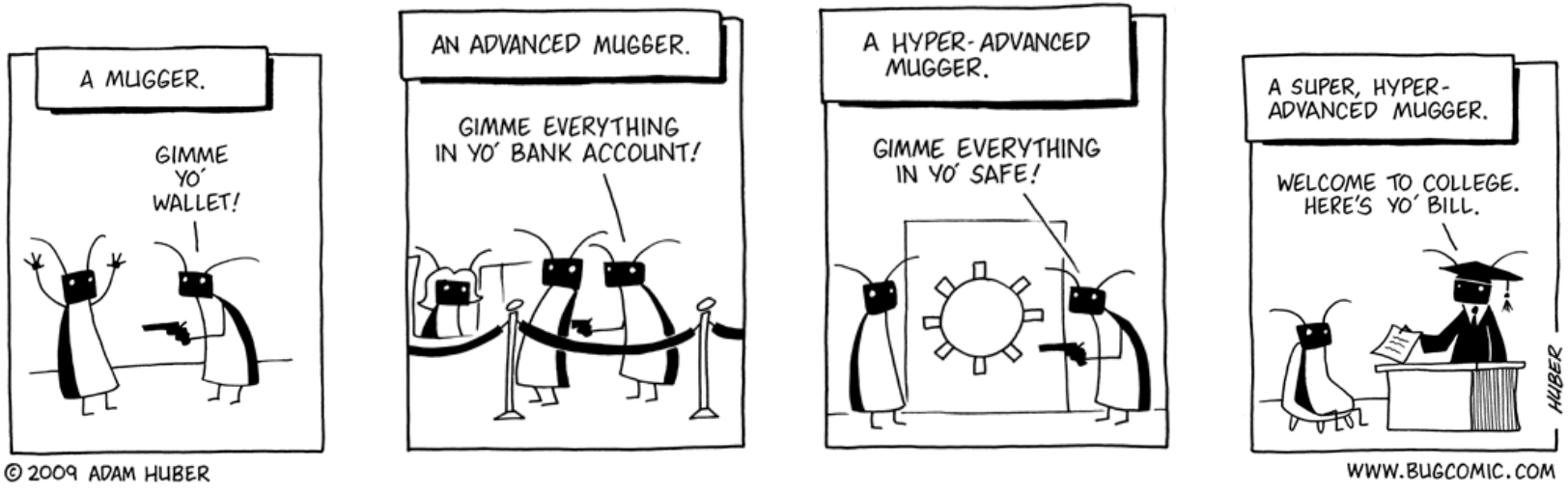
- Optional **Final project** for Python.
- Extra-credit, worth up to 20% of the total grade
- The due date for the final Python project will be announced closer to the end of the semester.

Soapbox: Thoughts on Grades



- Students are caught up in the **grade bubble**:
  - If I get good grades I will get a job.
  - If I get bad grades I will fail in life.
- In the real world, nobody cares about your GPA.
- Don't focus on grades!
- Focus on mastering the skills you need to get the job you want.

Soapbox: Thoughts on Grades



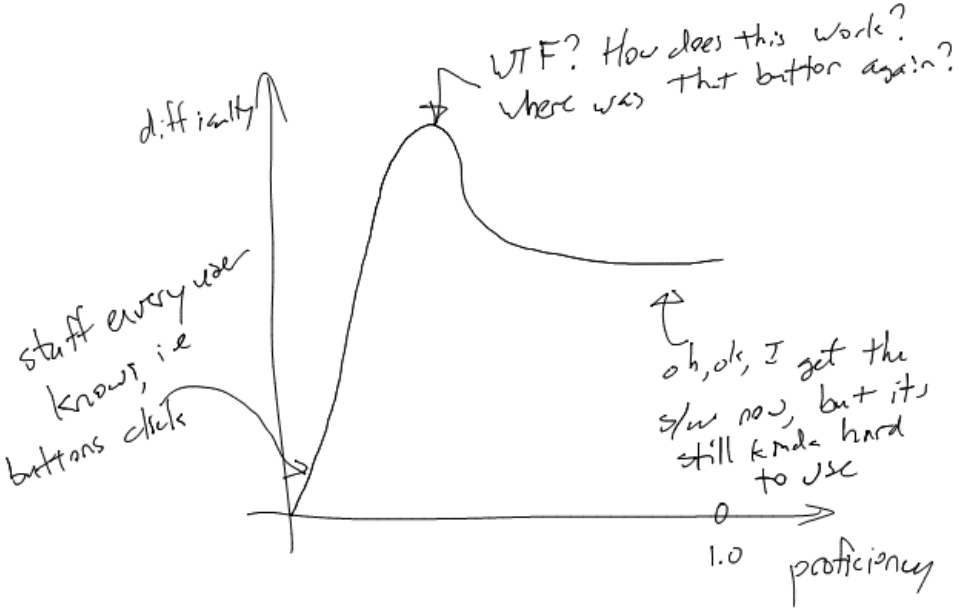
- Students are caught up in the **grade bubble**:
  - If I get good grades I will get a job.
  - If I get bad grades I will fail in life.
- In the real world, nobody cares about your GPA.
- Don't focus on grades!
- Focus on mastering the skills you need to get the job you want.

Soapbox: Our Role in This

- We are the academic equivalent of a **fitness trainer**.
- You're paying thousands of dollars for us to get you into shape, and we've developed a program for you that will do that.
- But you are the one who has to move their body.
- Bad techniques like cram learning may get you a good grade, but you're cheating yourself out of true progress.
- If you aren't working towards long-term intellectual growth, you're flushing tons of money down the toilet.

A Note on Python

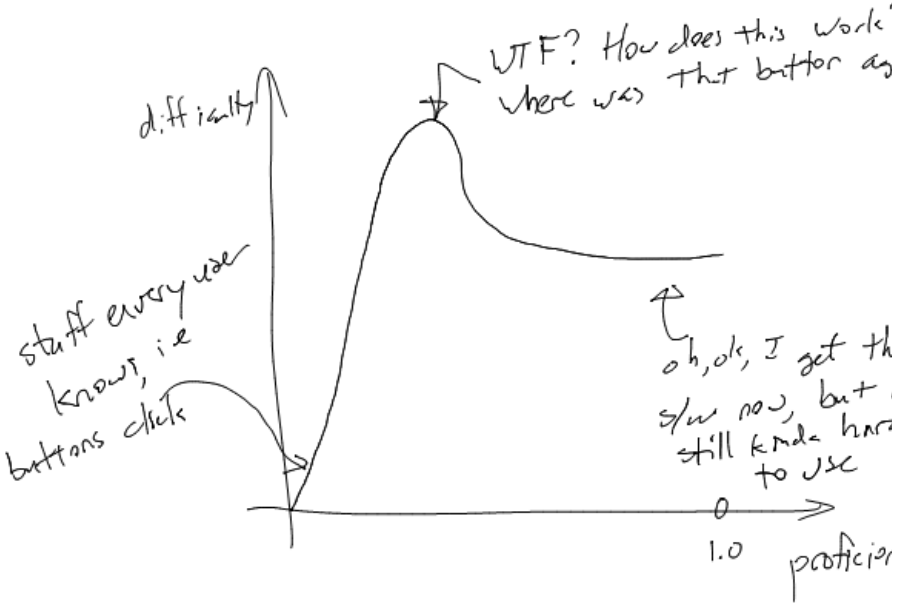
It's a steep learning curve!



A Note on Python

It's a steep learning curve!

- Don't despair! It takes time!



TL/DR: Ask for Help

- **Take advantage of us**
  We put a lot of effort into helping you achieve your goals:
  - recitations
  - office hours
  - availability via email and Google hangouts

  If you don't take advantage of these opportunities,
  you have no right to complain about grades or homeworks.
- **Take advantage of each other**
  Your peers are a valuable resource, too. Discuss homeworks, exchange ideas, share notes. Collaborate, help each other.
- **Don't wait too long**
  The Matthew effect also applies to education: the rich get richer, the poor get poorer. If you sense yourself falling behind, ask for help right away. The longer you wait, the worse it gets.

Getting Help

- By default: use discussion forums on CoCalc
- We have a team of Grad and Undergrad TAs: make use of them!
  - Their contacts and office hours are on Blackboard under the *Instructor Info* tab.

Optimizing Response Time

- Minor technical issue?
  → CoCalc Discussion Board > UGTA > TA > Instructors
- Minor/General Python Question?
  → CoCalc Discussion Board > UGTA > TA > Instructors
- Detailed Python/Homework question? → TA > Instructors
- Grading question? → TA > Instructors
- Talked to UG TAs/TAs but still have doubts? → Instructors
- Personal Issues? → Instructors

Getting Help [cont]

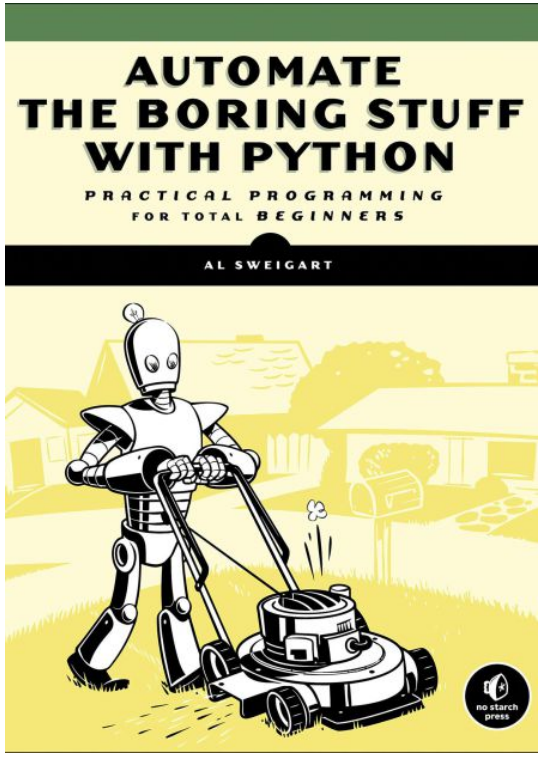
- Contacting us:
  - *{alena.aksenova,aniello.desanto}@stonybrook.edu*
  - Put [LIN120] at the beginning of the email *Object*
  - Reply time usually $<$ 24h (no guarantee during weekends!)
  - If you plan to come to our office hours, drop us a line the day before.
  - If there's a scheduling conflict, we'll let you know. Radio silence means everything is fine.
- For additional instructions, see the *Getting Help* section on Blackboard.

Some Final Remarks

1. **Course Website**
   - Familiarize yourself with CoCalc and Blackboard.
   - Lots of extra information there.
   - Check your SBU email frequently for Blackboard Announcements!
2. **Software Setup**
   - We will be using mostly **CoCalc**.
   - You will be invited to join today (via your SBU email).
   - You have to pay a 14 dollars subscription within the first two weeks, to ensure:
     - a fast CoCalc Virtual Machine
     - usable with internet connection
   - More information in Wednesday lecture and in the Friday recitation.
   - Get in touch if you have problems!

Supplementary Textbook (Optional!)

- Al Sweigart (2015):
  Automate the Boring Stuff with Python
- online version **free**
- digital versions and hardcopy around $25
- supplementary videos on Youtube
- **It is not required**
  but it's a good source to consult if something is unclear.



Tentative Schedule

	Theory	Python
Week 1	Syllabus	CoCalc & Notebook Tutorial
Week 2	Overview	Python Basics
Week 3	Overview	Strings
Week 4	Dialogue Systems	Control Flow
Week 5	Dialogue Systems	Lists & Loops
Week 6	Word Based Models	Summary & Practice
Week 7	More Word Based Models	Practice/Midterm
Week 8	**Spring Break**	**Spring Break**
Week 9	String Matching	String Cleaning
Week 10	N-gram Models	Functions
Week 11	Towards modern approaches: Neural Networks	Tokenizing
Week 12	Neural Networks & Deep Learning	Ngrams
Week 13	Human-Like Models?	Frequencies
Week 14	Summary: Impact on Society	Practice/Optional Final Project

Your First Homework

1. Carefully re-read this syllabus.
2. Read the document *How to Ace This Class* (it's on Blackboard).
3. Create a CoCalc account and play around with ti.

Note: There will be no mini-quiz next Monday.
Hint: But in general, each week check the quiz pool online (**on CoCalc**) for example questions.

Disability Support Services

If you have a physical, psychological, medical or learning disability that may impact your course work, please contact Disability Support Services, ECC (Educational Communications Center) Building, Room 128, (631) 632-6748. They will determine with you what accommodations, if any, are necessary and appropriate. All information and documentation is confidential.

Students who require assistance during emergency evacuation are encouraged to discuss their needs with their professors and Disability Support Services. For procedures and information go to the following website:

`http://www.stonybrook.edu/ehs/fire/disabilities`

Academic Integrity

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty are required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Technology & Management, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. For more comprehensive information on academic integrity, including categories of academic dishonesty, please refer to the academic judiciary website at `http://www.stonybrook.edu/uaa/academicjudiciary/`

Critical Incident Management

Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of Judicial Affairs any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn. Faculty in the HSC Schools and the School of Medicine are required to follow their school-specific procedures.