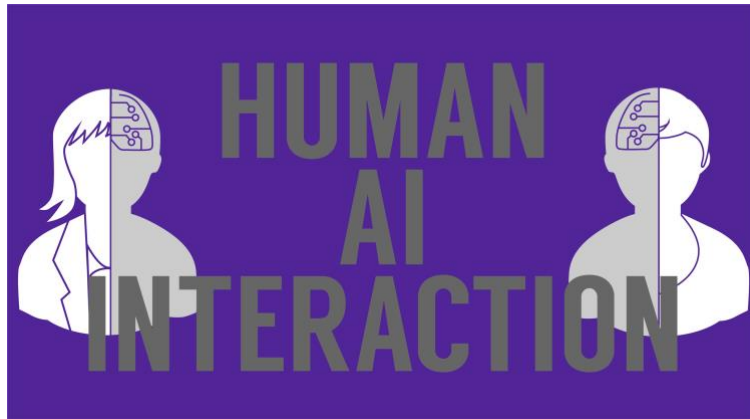




CSCI378: Human Artificial Intelligence Interaction (HAI)

<https://glow.williams.edu/courses/3876281>

Instructor:

Professor Iris Howley (she/her): call me "Iris" or "Professor Howley"

Email: ikh1@williams.edu

Student Help Hours: TBD (probably Wednesday afternoons...)

Course description:

As future innovators of the Artificial Intelligence frontier, we must be intentional about the impact our creations (and our creations' creations ...) have on our communities, our families, ourselves. AI is only useful in how it positively impacts the human experience, and in this course we explore how to harness the power of AI for the benefit of our world.

Course Structure

Learning Objectives:

After completing this course, students will be able to:

- Build intelligent interactive technologies
- Critically anticipate AI technology outcomes
- Intentionally design AI technologies for use by people
- Engage in current ongoing discussions of HAI

Communication:

My goal every semester is to support you in doing the best work you can in light of the challenges you face. I understand that Williams students face tremendous pressure to work hard, get “good” grades, and be as “successful” as possible. That said, I encourage you to remember that your

health and well-being are far more important than the work you do in this class or any class. And I encourage you to take the time you need to care for yourself and for your loved ones.

If you tell me you're having trouble, I will not judge you or think less of you. I hope you'll extend me the same grace. I also ask that you be patient with me if the challenges of this semester force me to make last-minute changes to the course plan. I will do my best to communicate any changes clearly and make them with respect for the inconvenience, frustration, and confusion that change may cause.

You *never* owe me personal information about your health (mental or physical). You are *always* welcome to talk to me about things that you're going through, though. If I can't help you, I usually know somebody who can. If you need extra help, or if you need more time with something, or if you feel like you're behind or not understanding everything, **do not suffer in silence!** Talk to me! I will work with you. **I promise.**

Please attend Student Help Hours. I'm also available through e-mail and the course discussion forum.

I want you to learn lots of things from this class, but I primarily want you to stay healthy, balanced and grounded.

Course Format:

This course places hands-on, active learning at its center and I am excited to provide a learning experience that fully enables these activities. As such, this course will be offered as a combination of asynchronous course materials (content delivery) with bi-weekly classes where we do small group activities and discussions. All course materials will be posted on and accessible through Glow. Deadlines will be available in Glow, but in general, prior to your Tuesday class, you should complete all readings, videos & listening materials, and concept review assignments for that week's module. Friday class will be for open discussion (lead by student pairs) and lab hours. I will generally try to post readings, videos & listening materials on Thursdays or Fridays. There will be approximately 4 larger assignments, and one final project as well.

Course Grading and Assignments

Grading:

A grading breakdown is below:

Attendance & Active Participation	15%
Discussion Leading	5%
Concept Reviews	20%
Assignments & Project	60%

Self-Reflections:

Concept Reviews are "ungraded." Your grade for the concept review will be determined based on your own self-assessment of your learning and effort on the readings, lectures, and other content

delivery, with the possibility of adjustments up or down from me. I reserve the right to alter your proposed concept review grade as appropriate, based on my own evaluation of your progress. If such an alteration seems warranted, I will contact you to set up a meeting to discuss your work in the course.

During the semester, you will be asked to complete three short self-reflection questionnaires. These questionnaires will include a series of questions about your work in the course. Specifically, you will be asked to:

- Talk about the work you have done in the course. e.g., How much effort did you commit to the reading and listening materials, the engagement activities, the concept reviews, the class videos, and the class projects?
- Talk about your successes and struggles in the course. e.g., When did you feel most successful in the class? What challenges did you encounter with the concepts, the assignments, or the logistics of the course?
- Talk about opportunities for improvement. e.g., What steps could you take to deepen your understanding of course concepts or increase your engagement in the class?
- Propose the grade you feel you should receive on your comprehension reviews (up to that point), based on your effort and comprehension of the readings, lectures, and other content delivery.
- Briefly explain why you would give yourself that grade.

Videos and Listening Materials:

Each week I will assign instructional videos, video lectures, or podcasts that must be watched prior to your first class meeting of the week. Some of these videos explain key AI or Human-Computer Interaction (HCI) concepts. Other videos offer examples of key concepts or explain how the reading materials illustrate these concepts. Still other videos introduce and explain the engagement activities and other assignments for the class. In general, there will be two sets of Video/Audio materials for each module. They are typically posted on Thursday or Friday.

Readings:

Each week includes one set of readings. These materials include a mix of news articles and academic journal articles. I have chosen these materials to illustrate key course concepts that I will be discussing in the recorded videos or that we will discuss in our class.

Concept Reviews:

Each week will include a short quiz to evaluate your comprehension of the materials. The concept reviews are required but I will not record your grades. Instead, I encourage you to use the reviews as an opportunity to evaluate your own understanding of the course material. Concept Reviews will typically be due prior to your first class session of the week.

Course Assignments:

You will be asked to complete 4 assignments and 1 project for this course. These assignments allow you to activate and demonstrate your understanding of the course concepts in creative and meaningful ways. Assignment deadline will also be preceded by a pass/fail Assignment Check-in that will help you distribute the effort required for the assignment. I will describe these assignments in more detail as we approach their release.

Discussion Leading:

Once per week, a pair of students will lead a discussion of that week's readings. An outline of the planned discussion will be due one day before the discussion.

Late Policy:

It is imperative that you complete the readings and watchings prior to class, otherwise you will not be able to participate. On all other deliverables for the course, three late days will be allowed over the entire semester no explanations needed, provided you email me prior to the assignment deadline letting me know your assignment will be late and how many late days you require. Additional late days are possible, but you must reach out to Iris prior to the deadline with an explanation of the situation.

Expected Workload

At Williams, we operate under the course unit system (rather than the credit hour system) as the metric required by many employers, granting organizations, graduate schools, and federal agencies. In addition to the 3 hours we spend together during our bi-weekly class time, you should expect to spend (on average) at least 12 hours per week on the academic and creative work related to class. If you find that you are spending considerably more (or considerably less!) time to engage with this course academically, please contact me so that we can determine the best course of action as you approach the materials. Should you have any additional questions about the relationship of course units to credit hours, refer to the Office of the Registrar, which [explains our course unit equivalency in greater detail](#).

Course Policies and Expectations

Learning Environment and Expectations:

Each person comes to the class with a unique background and perspective. I encourage you to draw on those perspectives when engaging with course materials and discussing them with your classmates. Please keep in mind that sharing opinions and experiences is a valuable but sometimes uncomfortable experience. Everyone must make the commitment to create an atmosphere of respect for each person's contribution. Varying points of view are welcome and expected. Please be respectful and open-minded when listening to viewpoints different from your own. If you disagree with an argument, criticize the evidence that supports a stance or the negative implications of a viewpoint; ask questions that challenge certain assumptions; but, please do not criticize the individual who holds that view. I do not tolerate discrimination.

Digital Devices:

You will need an internet-connected digital device that can run python and you can install additional python modules onto (i.e., a laptop or desktop computer), particularly for the course assignments. In particular, [Anaconda python](#) is a useful python distribution that comes with a variety of libraries that are necessary for this class.

I recognize that some students are unable to afford the cost of purchasing digital devices and that other students rely on older, more problem-prone devices that frequently break down or become unusable. I also recognize that those technology problems can be a significant source of stress for students.

If you do not have access to reliable internet or a reliable, internet-connected digital device for programming, I encourage you to contact the [Dean of Students Office](#) to request assistance. If you contact me, as well, I can help you write a message to the Dean of Students and also work with you to develop a plan for completing work in this course.

If you are experiencing problems with your device or internet access, I encourage you to contact the [Office of Information Technology for support](#).

If those problems are persistent and/or interfere with your ability to complete the work for this course, please let me know.

Please post a link to your favorite song about technology to the 'Random' discussion forum in Glow so I can see who is reading the syllabus :)

I also encourage you to be aware of the many technology-related resources that Williams College provides:

- Free on-campus [wireless internet](#) (wifi) accessed through the “eduroam” network. (For help connecting your device to the network, visit the [OIT website](#)).
- Free software (including Microsoft Office, Adobe Creative Suite, statistical software, etc.) described on the [OIT website](#).
- Free large quantity, secure online storage through [Google Drive](#) (a great way to back up files, [Drive File Stream](#) is particularly handy for accessing your Google Drive materials offline).
- Free regular [support](#) with issues related to OIT technology (e.g., email, Glow, wifi, printing, device setup, etc.).

Basic Needs Security:

Some Williams students experience difficulties affording groceries or accessing sufficient food to eat every day. Some lack a safe and stable place to live. Some Williams students experience challenges balancing school work with paid work, childcare, and other family obligations. Those challenges can create considerable anxiety for students and may also affect students' performance in their courses. Students experiencing those challenges are urged to contact the Dean of Students for support (online at <https://dean.williams.edu/> or by email at chaley@williams.edu or by phone at 413.597.4171). If you are experiencing challenges with food, housing, work, and/or family obligations, please notify me (ikh1@williams.edu) if you are comfortable in doing so. This will enable me to assist you in accessing support.

Family Commitments:

If you have care responsibilities for children or other family members, and if your caregiving responsibilities come into conflict with the course schedule, please don't feel as though you have to miss class. I understand that sometimes plans fall through. If this happens, you are welcome to contact me, and we can reschedule assignments.

Students Who Need Accommodations:

If formal accommodations need to be made to meet your specific learning or physical abilities, please contact me as soon as possible to discuss appropriate accommodations. Please also contact the Director of Accessible Education, Dr. G. L. Wallace (413.597.4672) or the Dean's office (413.597.4171). We will work together to ensure this class is as accessible and inclusive as possible. Also, students experiencing mental or physical health challenges that are significantly

affecting their academic work are encouraged to contact me or to speak with a dean. The deans can be reached at 413.597.4171.

Support for Students Experiencing Bias, Discrimination, Harassment or Abuse:

As your instructor, one of my responsibilities is to create a positive learning environment for all students. Title IX and Williams's Sexual Misconduct Policy prohibit sexual misconduct in any form, including sexual harassment, sexual assault, stalking, and dating and domestic violence. If you have experienced sexual misconduct, or know someone who has, the College can help. If you are seeking help and would like to speak to someone confidentially, you can contact: Integrative Wellbeing Services staff at 413.597.2353 (counseling services), The Rape and Sexual Assault Network ([RASAN](#)) at 413.597.4100 (peer advice services), Williams Health Center staff at 413.597.2206 (health and medical services). It is also important that you know that Title IX and College policy require professors to share any information brought to their attention about potential sexual misconduct, with the Williams's Title IX Coordinator. In that event, those individuals will work to ensure that appropriate measures are taken and resources are made available. Protecting student privacy is of utmost concern, and information will only be shared with those that need to know to ensure the College can respond and assist. We encourage you to visit <https://titleix.williams.edu/for-students/> to learn more.

More generally, any act of discrimination or harassment based on race, ethnicity, religious affiliation, gender, gender identity, sexual orientation or disability can be reported through any of these options: 1) complete the Office of Institutional Diversity, Equity, and Inclusion's ([OIDEI](#)) [online form](#); 2) call the Office of Institutional Diversity, Equity, and Inclusion at 413.597.4376; or 3) contact the Assistant Vice President for Institutional Diversity and Equity/Title IX Coordinator.

Academic Integrity:

Assignments are to be the sole work of each student unless the assignment explicitly states otherwise. Students may discuss issues related to an assignment, provided that such discussions are cited in the material turned in. Any ideas, code, images, or text that is not original to the student(s) must have the original source cited, both in the code comments and in the display text. Uncredited collaborations will be considered a violation of the honor code and will be handled appropriately. For a full description of the Computer Science Honor Code, please see <https://csci.williams.edu/the-cs-honor-code-and-computer-usage-policy/> If in doubt of what is appropriate, do not hesitate to ask.

Furthermore, part of your grade in this class relies on your self-reflection of your own progress, it is imperative that you are honest in your assessment of your activities for this class. i.e., it is a breach of Honor Code to falsely claim completed work that you have not done for this course.

Intellectual Property:

No part of this course may be reproduced nor distributed in any manner without prior permission from the instructor.

Course Content

The content of this course will be accessible through materials posted on the Glow site. In order to be prepared for weekly classes, I encourage you to complete each assignment by the

recommended completion date listed with that assignment. If you experience problems that delay your progress through these assignments, please let me know.

The content of the topics is subject to change. However, the order of the topics will, roughly, be as follows:

Module 0: Getting Started in the Class

Why does Human-AI interaction matter? Who is this Iris person? What are the basic course logistics? What social distancing related things should I be aware of?

Module 1: Introduction to Human-AI Interaction

What is AI? What is machine learning? What is the machine learning pipeline?

Module 2: Perspectives on Human-AI Interaction

Why were there previous "AI winters"? What is the history of humans interacting with AI? Artificial Intelligence versus Intelligence Augmentation.

Module 3: Designing AI/ML User Experiences

How do we take into consideration user needs and values as we design user-facing AI systems?

Module 4: Designing for Failure

How do we design around the inevitable AI failures for our stakeholders?

Module 5: Data & Knowledge

Where does data come from and who owns it? How do we get enough data (ethically) for our AI systems?

Module 6: Visualizing Complex Predictive Data

Designing visualizations to improve user understanding of AI outputs, as well as designing visualizations to assist developers who build AI/ML systems.

Module 7: Interpreting and Explaining Algorithms

How do we explain our AI systems for sufficient user understanding so users can interrogate the system's predictions?

Module 8: AI Ethics, Fairness, Social Acceptability, and Trust

Applying Distributional justice approaches to improve fairness in AI systems.

Module 9: Humans in the Loop

Building systems that leverage collaboration between humans and AI systems.

Module 10: Natural Language and Speech Applications

Using neural nets for conversational agents.

Module 11: Vision, Images, and Art + AI Jobs

The pros, cons, and how-tos of AI-generated imagery, GANs.

Module 12: Recommender Systems + Course Wrap-up

Recommender Systems, continuing to study HAll, course wrap-up

Recent Announcements



Today's Project Pitch Slides

https://glow.williams.edu/courses/3876281/discussion_topics/23547534

Hi folks -In case you're interested, I've uploaded today's Project Pitch slide deck here. We're going for a reverse alphabetical order by last name. See you very soon!- Iris Howley

[← Reply](#)

https://glow.williams.edu/courses/3876281/discussion_topics/23547534

Posted on:

May 10, 2024, 12:41 PM



Module 11 - Discussion Outline Available

https://glow.williams.edu/courses/3876281/discussion_topics/23520291

Hi folks! Last Reading Discussion of the semester is tomorrow! The Discussion Outline from your classmates is available here.- Iris Howley

[← Reply](#)

https://glow.williams.edu/courses/3876281/discussion_topics/23520291

Posted on:

May 2, 2024, 6:28 PM



Discussion Outline for tomorrow/NLP available!

https://glow.williams.edu/courses/3876281/discussion_topics/23496872

Hi folks -Same old, same old. Find the Module 10 Discussion Outline here.- Iris Howley

[← Reply](#)

https://glow.williams.edu/courses/3876281/discussion_topics/23496872

Posted on:

Apr 25, 2024, 3:15 PM

Collapse All

[View Progress](#)

▼ #0 Course Information (Getting Started in the Course)		Complete All Items
		⋮
	Task List for Module #0 - starting 2/2	⋮
	Attend Your Tues/Fri Class	⋮
	Lecture 1: Welcome!  (https://www.youtube.com/watch?v=X82cvUv7nbY&list=PLMfRT7Ik6baF1TqU4v5Bcec6FmSC55ttJ)	
	Lecture00-1_CourseIntro.pdf	⋮
	Start of Semester Survey Feb 4, 2024 1 pts View	⋮
	Read: How to Read & Understand a Scientific Paper  (https://www.huffpost.com/entry/how-to-read-and-understand-a-scientific-paper_b_5501628?guccounter=1&guce_referrer=aHR0cHM6Ly9kdWNrZHVja2dvLmNvbS8&guce_referrer_sig=AQAAALmPclpXfi0Pp9YPZ6sdSCTz0WfhSn-1W6RwzDIccQBvtnpM6b9KBQETfmBQrhzg6Pk2EZhVOkcGu9vj4Vzkdnf_vWkrn3IGkX7BeTJI0ZvtobWxJORp3cXlah3_qtIMHs-ycA0yGdAB_CTbRy3vrkunBZapGg3qdhtyFDu1Kncm)	
	CSCI378syllabus-s24.pdf	⋮
	Student Help Hours	⋮
	Syllabus Review Feb 6, 2024 4 pts	⋮

#1 Practical Machine Learning

Complete All Items

⋮



Task List for Module #1 - starting 2/2

⋮



Read: Direct Manipulation vs. Interface Agents

View

⋮



[Optional Reading: The Coming AI Autumn](#) 

https://jeffreybigham.com/blog/2019/the-coming-ai-autumnn.html#ftnt_ref1



[Lecture 1: Practical ML](#) 

https://www.youtube.com/watch?v=HVG8IJ2PzmE&list=PLMfRT7Ik6baHufQulhMI21WABc_zjUFx2



Lecture01-1_PracticalML.pdf

⋮



Attend Your Tues/Fri Class

⋮



Assignment 1: ML for Home Loans

Feb 21, 2024 4 pts

⋮



Assignment 1 Check-in

Feb 14, 2024 1 pts

⋮

#2 Perspectives on HAI

Complete All Items

⋮



Task List for Module #2 - starting 2/9

⋮



Read: Principles of Mixed-Initiative Interaction

View

⋮

	Read: Man-Computer Symbiosis	
	Lecture 1: History of AI  (https://www.youtube.com/watch?v=xCp2CxEs3w0&list=PLMfRT7Ik6baHmOXppBG9A1k2VYtT9trUa)	
	Lecture02-1_HistoryOfAI.pdf	
	Optional Podcast: Judge John Hodgeman - "Are machine guns robots?" 11/15/2010  (https://open.spotify.com/episode/7j26AVd5Hpf2qHp8l3XHkx?si=QJfTwuM-QpGSJVHgLV97w)	
	Lecture 2: History of Humans Interacting w AI (Intelligence Augmentation)  (https://www.youtube.com/watch?v=Yt70KG36-DQ&list=PLMfRT7Ik6baGqoCFP4DZVtvhMkOM8BxKI)	
	Lecture02-2_AlviA.pdf	
	Module #2 Concept Review Feb 13, 2024 1 pts	
	Attend Your Tues/Fri Class	
	Assignment 1: ML for Home Loans Feb 21, 2024 4 pts	
	Assignment 1 Check-in Feb 14, 2024 1 pts	

▼ #3 Designing AI/ML User Experiences

Complete All Items



Task List for Module #3 - starting 2/16



 [Read: The Designer's Guide to Machine Learning](https://digitalist.global/talks/the-designers-guide-to-machine-learning/)  (https://digitalist.global/talks/the-designers-guide-to-machine-learning/)

 **Read: The Designer's Guide to Machine Learning** 

 **Read: Guidelines for Human-AI Interaction** 
View

 [Read: Software engineering for machine learning: a case study \[blog post\]](https://blog.acolyer.org/2019/07/08/software-engineering-for-machine-learning/)  (https://blog.acolyer.org/2019/07/08/software-engineering-for-machine-learning/)

 **Optional Reading: Software engineering for machine learning: a case study [full article]** 

 **Optional Reading: How Good is 85%? A Survey Tool to Connect Classifier Evaluation to Acceptability of Accuracy** 

 [Lecture 1: Matchmaking Needs & Risks for Adding AI/ML](https://www.youtube.com/watch?v=M6Df9lycMNM&list=PLMfRT7Ik6baGuHcrNWMt8jT5rQhSGrbqI)  (https://www.youtube.com/watch?v=M6Df9lycMNM&list=PLMfRT7Ik6baGuHcrNWMt8jT5rQhSGrbqI)

 **Lecture03-1_MatchmakingAIML.pdf** 

 **Module 3.1 Worksheets**

 [Lecture 2: Checklists and Principles and Values](https://www.radicalai.org/michael-madaio)  (https://www.radicalai.org/michael-madaio)

 [Optional Lecture: Google I/O '19 Designing Human-Centered AI Products](https://www.youtube.com/watch?v=rf83vRxLWFQ)  (https://www.youtube.com/watch?v=rf83vRxLWFQ)

	Module #3 Concept Review Feb 20, 2024 1 pts	⋮
	Attend Your Tues/Fri Class	⋮
	Assignment 1: ML for Home Loans Feb 21, 2024 4 pts	⋮

▼ #4 Designing for Failure		Complete All Items ⋮
	Task List for Module #4 - starting 2/23	⋮
	Read: UX designers pushing AI in the enterprise: A case for adaptive UIs	⋮
	Read: UX designers pushing AI in the enterprise: A case for adaptive UIs  (https://interactions.acm.org/archive/view/january-february-2021/ux-designers-pushing-ai-in-the-enterprise)	
	Read: Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-User Expectations of AI Systems View	⋮
	Read: Can we make Montreal's buses more predictable? No. But machines can  (https://blog.transitapp.com/can-we-make-montreals-buses-more-predictable-no-but-machines-can-e42f28a1a0ba/)	
	Optional Reading: "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making	⋮

 [Lecture 1: Failure & Feedback with Users](https://www.youtube.com/watch?v=_0ECHj-1fig&list=PLMfRT7Ik6baG2FcWNEqA5gdz4Zq62cBHR)  [_0ECHj-1fig&list=PLMfRT7Ik6baG2FcWNEqA5gdz4Zq62cBHR](https://www.youtube.com/watch?v=_0ECHj-1fig&list=PLMfRT7Ik6baG2FcWNEqA5gdz4Zq62cBHR)

 Lecture04-1_AIMLfailUX.pdf

 [Lecture 2: Avoiding AI/ML Failures with People](https://www.youtube.com/watch?v=deDiGrJD32g&list=PLMfRT7Ik6baHbNScvsbGBaUALe-OHhMsj)  [_deDiGrJD32g&list=PLMfRT7Ik6baHbNScvsbGBaUALe-OHhMsj](https://www.youtube.com/watch?v=deDiGrJD32g&list=PLMfRT7Ik6baHbNScvsbGBaUALe-OHhMsj)

 Lecture04-2_AIMLfailCollection.pdf

 Optional Lecture: Joy Buolamwini's "AI, Ain't I a Woman?"

 **Module #4 Concept Review**
Feb 27, 2024 1 pts

 **Attend Your Tues/Fri Class**

 **Assignment 2: Designing AI/ML Front Ends**
Mar 15, 2024 4 pts

 **Assignment 2 Check-in**
Mar 6, 2024 1 pts

▼ **#5 Data & Knowledge: Where does it come from and who owns it?**

 **Task List for Module #5 - starting 3/1**

 **Read: The Trust Imperative, a Framework for Ethical Data Use**

Read: Captcha if you can: how you've been training AI for years without realising it



[_](https://www.techradar.com/news/captcha-if-you-can-how-youve-been-training-ai-for-years-without-realising-it)(<https://www.techradar.com/news/captcha-if-you-can-how-youve-been-training-ai-for-years-without-realising-it>)



Read: Why Google Maps Is Better Than Apple Maps



Optional: People Have Begun to Love Apple's Most Hated Product [_](https://www.wsj.com/articles/apple-maps-app-popularity-iphone-8e52aec1)(<https://www.wsj.com/articles/apple-maps-app-popularity-iphone-8e52aec1>)



Read: Human-Computer Interaction and Collective Intelligence



Read: Underpaid Workers Are Being Forced to Train Biased AI on Mechanical Turk
[_](https://www.vice.com/en/article/88apnv/underpaid-workers-are-being-forced-to-train-biased-ai-on-mechanical-turk)(<https://www.vice.com/en/article/88apnv/underpaid-workers-are-being-forced-to-train-biased-ai-on-mechanical-turk>)



Optional Reading: If Big Tech has our data, why are targeted ads so terrible?
[_](https://www.ft.com/content/b013d9a2-c69d-4c17-aaeb-020eb2e33403)(<https://www.ft.com/content/b013d9a2-c69d-4c17-aaeb-020eb2e33403>)



Lecture 1: AI Design & Data [_](https://www.youtube.com/watch?v=hrXlhvg5_FA&list=PLMfRT7Ik6baH5WYWZH2r3O-T5iXo8gnMv)(https://www.youtube.com/watch?v=hrXlhvg5_FA&list=PLMfRT7Ik6baH5WYWZH2r3O-T5iXo8gnMv)



Lecture05-1_IntroDataForAIdesign.pdf



Optional Podcast: Data as Protest: Data for BLM



Lecture 2: People & Public Data for AI [_](https://www.youtube.com/watch?v=9hZ2swx7vVs&list=PLMfRT7Ik6baEXHjot-7i3S_1ZK2OyncRx)(https://www.youtube.com/watch?v=9hZ2swx7vVs&list=PLMfRT7Ik6baEXHjot-7i3S_1ZK2OyncRx)



Lecture05-2_PeoplePublicDataAI.pdf



Lecture 3: Collecting human data for AI/ML  (https://www.youtube.com/watch?v=rFltCw6oRPw&list=PLMfRT7Ik6baFlem_3WR5P7kNvafuJJEmu)



 **Lecture05-3_CollectingDataForAIML.pdf** 



Optional Lecture: Reinforcement learning in Minecraft: Challenges and opportunities in multiplayer games  (https://note.microsoft.com/MSR-Webinar-Project-Malmo-Registration-On-Demand.html?wt.mc_id=email_MSR-WBNR_first-invite_2021-02-18)



Module #5 Concept Review 
Mar 5, 2024 1 pts



Attend Your Tues/Fri Class 



Assignment 2 Check-in 
Mar 6, 2024 1 pts



Assignment 2: Designing AI/ML Front Ends 
Mar 15, 2024 4 pts

▼ **#6 Data Visualization & Communication**

Complete All Items



Task List for Module #6 - starting 3/8 



Mid-Semester Feedback 
Mar 14, 2024



Read: Human-Centered Tools for Coping with Imperfect Algorithms during Medical Decision-Making 
View

	Read: When (ish) is My Bus? User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems	
	Optional Reading: Seq2Seq: Visual Debug Tool for Seq-to-Seq Models	
	Optional Reading: xAI Understanding, Visualizing, and Interpreting Deep Learning Models	
	Lecture 1: Visualization to improve HAIL  (https://www.youtube.com/watch?v=4vzbk4hdnNo&list=PLMfRT7Ik6baEua2C0cg0RNNRKuuXrGqiq)	
	Lecture06-1_VizToImproveHAIL.pdf	
	Optional Lecture: Data Visualization: Bridging the Gap Between Users and Information  (https://www.microsoft.com/en-us/research/video/data-visualization-bridging-the-gap-between-users-and-information/)	
	Optional Podcast: Visualizing Uncertainty with Jessica Hullman and Matthew Kay  (https://datastori.es/134-visualizing-uncertainty-with-jessica-hullman-and-matthew-kay/)	
	Lecture 2: Info Viz for Users of AI/ML Systems  (https://www.youtube.com/watch?v=uTv3TPPjtQY&list=PLMfRT7Ik6baGsr526y2Ssy6mxuTRLk-uw)	
	Lecture06-2_VizForAIML.pdf	
	Optional Podcast: Human-Driven Machine Learning with Saleema Amershi  (https://datastori.es/115-human-driven-machine-learning-with-saleema-amershi/)	
	Module #6 Concept Review Mar 12, 2024 1 pts	

	Attend Your Tues/Fri Class	⋮
	Assignment 2: Designing AI/ML Front Ends Mar 15, 2024 4 pts	⋮
	Assignment 3: Viz for AI Apr 17, 2024 4 pts	⋮
	Assignment 3 Check-in Apr 10, 2024 1 pts	⋮

▼ #7 Interpreting and Explaining Algorithms		Complete All Items ⋮
	Task List for Module #7 - starting 3/15	⋮
	Mid-Semester Feedback Mar 14, 2024	⋮
	Read: Challenge of Crafting Intelligible Intelligence View	⋮
	Read: Why these Explanations? Selecting Intelligibility Types for Explanation Goals	⋮
	Optional Reading: The Mythos of Model Interpretability	⋮
	Optional Reading: Why Should I Trust You? Explaining the Predictions of Any Classifier	⋮
	Lecture 1: Interpreting & Explaining Algorithms  (https://www.youtube.com/watch?v=HJe-7Vew-wE&list=PLMfRT7Ik6baGdz7ZZ7kwhnLeZfDgXTJXx)	



Lecture07-1_InterpretExplainAI.pdf



Optional Lecture: Jenn Wortman Vaughan's "Transparency and Intelligibility Throughout the Machine Learning Life Cycle"



[Lecture 2: Designing Explanations for AI](https://www.youtube.com/watch?v=SBMJQubASQE&list=PLMfRT7Ik6baGhvAhopWeqvVHulRh1ZAIy) (<https://www.youtube.com/watch?v=SBMJQubASQE&list=PLMfRT7Ik6baGhvAhopWeqvVHulRh1ZAIy>)



Lecture07-2_DesigningExplainAI.pdf



Module #7 Concept Review

Apr 2, 2024 1 pts



Attend Your Tues/Fri Class



Assignment 3: Viz for AI

Apr 17, 2024 4 pts



Assignment 3 Check-in

Apr 10, 2024 1 pts



▼ #8 AI Ethics, Fairness, and Trust



Task List for Module #8 - starting 4/5



[Read: Facial Recognition Is the Perfect Tool for Oppression](https://medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66) (<https://medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66>)



[Optional Reading: Your Face is Not Your Own](https://www.nytimes.com/interactive/2021/03/18/magazine/facial-recognition-clearview-) (<https://www.nytimes.com/interactive/2021/03/18/magazine/facial-recognition-clearview->

[ai.html?searchResultPosition=2\)](#)

Optional Reading: Time to regulate AI that interprets human emotions 

 (<https://www.nature.com/articles/d41586-021-00868-5>)

Read: How Facebook got addicted to spreading misinformation 

 (<https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/>)

Optional Reading: Facebook Doesn't Need a Chief Ethics Officer 

 (<https://slate.com/technology/2017/01/a-chief-ethics-officer-wont-fix-facebooks-problems.html>)

Optional Reading: Google might ask questions about AI ethics, but it doesn't want answers

 **Read: Disparate Interactions**

⋮

 **Optional Reading: Do artifacts have politics?**

⋮

 **Optional Reading: Residual Unfairness in Fair Machine Learning from Prejudiced Data**

⋮

 **Optional Reading: Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission**

⋮

 **Lecture 1: AI Ethics, Fairness, Bias**  (<https://www.youtube.com/watch?v=ZZiqK-B9u4Q&list=PLMfRT7Ik6baHtyGhNFlioXX9HJUtgAbUp>)

 **Lecture08-1_EthicsFairnessTrust.pdf**

⋮



[Optional Lecture: Alex Hanna - Data, Transparency, and AI Ethics \(Ethics of AI in Context\)](#)  <https://youtu.be/nBGEsBljIW0>



[Lecture 2: Kate Crawford, "The Trouble with Bias"](#)  <https://youtu.be/ggzWlipKraM>



[Optional Lecture: Ruha Benjamin - "Which Humans? Innovation, Equity, and Imagination in Human-Centered Design"](#)  <https://www.youtube.com/watch?v=kDcz44ifdQw>



Module #8 Concept Review

Apr 9, 2024 1 pts





Attend Your Tues/Fri Class





Assignment 3 Check-in

Apr 10, 2024 1 pts





Assignment 3: Viz for AI

Apr 17, 2024 4 pts





#9 Humans in the Loop with AI/ML





Task List for Module #9 - starting 4/12





[Read: Humans in the Loop: The Design of Interactive AI Systems](#)  <https://hai.stanford.edu/blog/humans-loop-design-interactive-ai-systems>



Read: Machine Learning with Humans in the Loop



	Read: Evorus: A Crowd-powered Conversational Assistant Built to Automate Itself Over Time	⋮
	Optional Reading: Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Tradeoff	⋮
	Optional Reading: In Search of the Dream Team: Temporally Constrained Multi-Armed Bandits for Identifying Effective Team Structures	⋮
	Lecture 1: Humans in the Loop for Data  (https://www.youtube.com/watch?v=ZrGyjAB0f3Y&list=PLMfRT7Ik6baEHtPoi_gZVWdlyzGKCwGzR)	
	Lecture09-1_HumansInLoopData.pdf	⋮
	Lecture 2: Humans in the Loop for Human-AI Partnerships  (https://www.youtube.com/watch?v=DQ9NMj5O8FQ&list=PLMfRT7Ik6baEEdNZFH0zhJ1wjxp1DjhQ9)	
	Lecture09-2_HumansInLoopPartnership.pdf	⋮
	Optional Lecture: Making fashion recommendations with human-in-the-loop machine learning  (https://www.youtube.com/watch?v=aTB38biOBoE)	
	Module #9 Concept Review Apr 16, 2024 1 pts	⋮
	Attend Your Tues/Fri Class	⋮
	Assignment 3: Viz for AI Apr 17, 2024 4 pts	⋮

Complete All Items

▼ #10 Natural Language & Conversational Agents

⋮



Task List for Module #10 - starting 4/19

⋮



Read: Mind: How to Build a Neural Network (Part One) ~ focus on the high-level process! [↗\(https://stevenmiller888.github.io/mind-how-to-build-a-neural-network/\)](https://stevenmiller888.github.io/mind-how-to-build-a-neural-network/)



Optional Lecture: Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) [↗\(https://www.youtube.com/watch?v=WCUNPb-5EYI\)](https://www.youtube.com/watch?v=WCUNPb-5EYI)



Read: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

⋮



Read: Beyond Dyadic Interactions: Considering Chatbots as Community Members

⋮



Optional Reading: "Hey Google is it OK if I eat you?" Initial Explorations in Child-Agent Interaction

⋮



Optional Reading: ELIZA—a computer program for the study of natural language communication between man and machine

⋮

[View](#)

Optional Reading: Voice Interfaces in Everyday Life

⋮



Optional Reading: Speak, Memory: When her best friend died, she rebuilt him using AI



Lecture 1: Natural Language & Conversational Agents [↗\(https://www.youtube.com/watch?v=RAqRVYf4Dq4&list=PLMfRT7Ik6baFLpF-ZwELwywpik9iwlaE2\)](https://www.youtube.com/watch?v=RAqRVYf4Dq4&list=PLMfRT7Ik6baFLpF-ZwELwywpik9iwlaE2)

	Lecture10-1_NaturalLangConvAgents.pdf	⋮
	Lecture 2: Radical AI's "The Power of Linguistics" w Emily Bender (can start at 13:23)  _ (https://www.youtube.com/watch?v=WWnrKRAMPng)	
	Optional Lecture: MSR's "Failures of imagination: Discovering and measuring harms in language technologies"	
	Module #10 Concept Review Apr 23, 2024 1 pts	⋮
	Attend Your Tues/Fri Class	⋮
	Project Check-in Apr 30, 2024 1 pts	⋮
	Assignment 4: Satire Chatbot May 2, 2024 4 pts	⋮
	Assignment 4 Check-in Apr 25, 2024 1 pts	⋮
	Final Project May 15, 2024 4 pts	⋮
	Project Pitch May 9, 2024 2 pts	⋮
▼ #11 Vision, Images, and Art		⋮
	Task List for Module #11 - starting 4/26	⋮

Read: Slick Tom Cruise Deepfakes Signal That Near Flawless Forgeries May Be Here

 [_ \(https://www.npr.org/2021/03/11/975849508/slick-tom-cruise-deepfakes-signal-that-near-flawless-forgeries-may-be-here\)](https://www.npr.org/2021/03/11/975849508/slick-tom-cruise-deepfakes-signal-that-near-flawless-forgeries-may-be-here)

Read: Using AI to Detect Seemingly Perfect Deep-Fake Videos

 [_ \(https://hai.stanford.edu/blog/using-ai-detect-seemingly-perfect-deep-fake-videos\)](https://hai.stanford.edu/blog/using-ai-detect-seemingly-perfect-deep-fake-videos)

Read: A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way

 [_ \(https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53\)](https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53)

Optional Reading: Image Kernels Explained Visually

 [_ \(https://setosa.io/ev/image-kernels/\)](https://setosa.io/ev/image-kernels/)

Read: Artificial Art: How GANs are making machines creative

 [_ \(https://heartbeat.fritz.ai/artificial-art-how-gans-are-making-machines-creative-b99105627198\)](https://heartbeat.fritz.ai/artificial-art-how-gans-are-making-machines-creative-b99105627198)

Optional Reading: Some cool applications of GAN

 [_ \(https://jonathan-hui.medium.com/gan-some-cool-applications-of-gans-4c9ecca35900\)](https://jonathan-hui.medium.com/gan-some-cool-applications-of-gans-4c9ecca35900)

Optional Reading: Google's ML Crash Course on GANs

 [_ \(https://developers.google.com/machine-learning/gan\)](https://developers.google.com/machine-learning/gan)

Lecture 1: AI & Images (Processing & Generation)

 [_ \(https://www.youtube.com/watch?v=2jB58iX3TIQ&list=PLMfRT7Ik6baHE4ScCiJvTBI7iWZa4CpHy\)](https://www.youtube.com/watch?v=2jB58iX3TIQ&list=PLMfRT7Ik6baHE4ScCiJvTBI7iWZa4CpHy)

 **Lecture11-1_VisionImagesArt.pdf**

⋮

Optional Lecture: How Convolutional Neural Networks work

 [_ \(https://www.youtube.com/watch?v=FmpDlaiMleA\)](https://www.youtube.com/watch?v=FmpDlaiMleA)

[Optional Lecture: A Friendly Introduction to GANs](#)

<https://www.youtube.com/watch?v=8L11aMN5KY8>

[Lecture 2: Computer vision in practice: who is benefiting and who is being harmed?](#)

<https://youtu.be/0sBE5OyD7fk>

Lecture11-2_FATEinCV.pdf

[Optional Lectures: Part 2 & 3 in the FATE/CV Workshop](#)

<https://sites.google.com/view/fatecv-tutorial/schedule>

Module #11 Concept Review

Apr 30, 2024 1 pts

Attend Your Tues/Fri Class

Final Project

May 15, 2024 4 pts

Project Check-in

Apr 30, 2024 1 pts

Project Pitch

May 9, 2024 2 pts

#12 Recommender Systems & AI in the World

Task List for Module #12 - starting 5/3

[Read: YouTube, the Great Radicalizer](#)

<https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>

 [Optional Reading: Extremism pays. That's why Silicon Valley isn't shutting it down](https://www.theguardian.com/commentisfree/2018/mar/18/extremism-pays-why-silicon-valley-not-shutting-it-down-youtube)  (<https://www.theguardian.com/commentisfree/2018/mar/18/extremism-pays-why-silicon-valley-not-shutting-it-down-youtube>)

 **Optional Reading: Deep Neural Networks for YouTube Recommendations** 

 **Optional Reading: Exploring the Filter Bubble: Effect of using recommender systems on content diversity** 

 **Optional Reading: Understanding the role of latent feature diversification on choice difficulty and satisfaction** 

 **Read: Fix Section 230 and Hold Tech Companies to Account**  (<https://www.wired.co.uk/article/section-230-social-media>)

 **Optional Reading: The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People** (<https://glow.williams.edu/courses/3490440/files/folder/reading?preview=249876273>)

 **Optional Reading: Congress is way behind on algorithmic misinformation**  (<https://www.theverge.com/2021/4/27/22406054/facebook-twitter-google-youtube-algorithm-transparency-regulation-misinformation-disinformation>)

 **Read: Google is Poisoning its Reputation with AI Researchers**  (<https://www.theverge.com/2021/4/13/22370158/google-ai-ethics-timnit-gebru-margaret-mitchell-firing-reputation>)

 **Optional Reading: Google Plans to Double Its AI Ethics Research Staff**  (<https://www.wsj.com/articles/google-plans-to-double-ai-ethics-research-staff-11620749048>)

[Optional Reading: "What on earth are they doing?": AI ethics experts react to Google doubling embattled ethics team](https://www.morningbrew.com/emerging-tech/stories/2021/05/14/earth-ai-ethics-experts-react-google-doubling-embattled-ethics-team)



<https://www.morningbrew.com/emerging-tech/stories/2021/05/14/earth-ai-ethics-experts-react-google-doubling-embattled-ethics-team>

 [Lecture 1: Recommender Systems](https://www.youtube.com/watch?v=_yDcxfBjgQ8&list=PLMfRT7lk6baHhBjhD6M0exQ1CcTFIk5Y)  https://www.youtube.com/watch?v=_yDcxfBjgQ8&list=PLMfRT7lk6baHhBjhD6M0exQ1CcTFIk5Y



Lecture12-1_RecommenderSystems.pdf



 [Optional Podcast: NYTimes "Rabbit Hole, What is the internet doing to us?"](https://www.nytimes.com/column/rabbit-hole) 

<https://www.nytimes.com/column/rabbit-hole>

 [Optional Podcast: Radical AI "What is AI for social good?"](https://www.radicalai.org/ai-for-social-good-panel) 

<https://www.radicalai.org/ai-for-social-good-panel>

 [Lecture 2: AI & the World + Course Wrap-up](https://www.youtube.com/watch?v=SKS2QHieNPU&list=PLMfRT7lk6baE9Gw4OApwKTQwxut8LFZhK)  <https://www.youtube.com/watch?v=SKS2QHieNPU&list=PLMfRT7lk6baE9Gw4OApwKTQwxut8LFZhK>



Lecture12-2_AlandJobs.pdf



 [Optional Podcast: Radical AI "Industry AI Ethics 101 with Kathy Baxter"](https://www.radicalai.org/industry-ai-ethics) 

<https://www.radicalai.org/industry-ai-ethics>



Module #12 Concept Review

May 7, 2024 1 pts



Project Check-in

Apr 30, 2024 1 pts



Project Pitch

May 9, 2024 2 pts



Final Project

May 15, 2024 4 pts

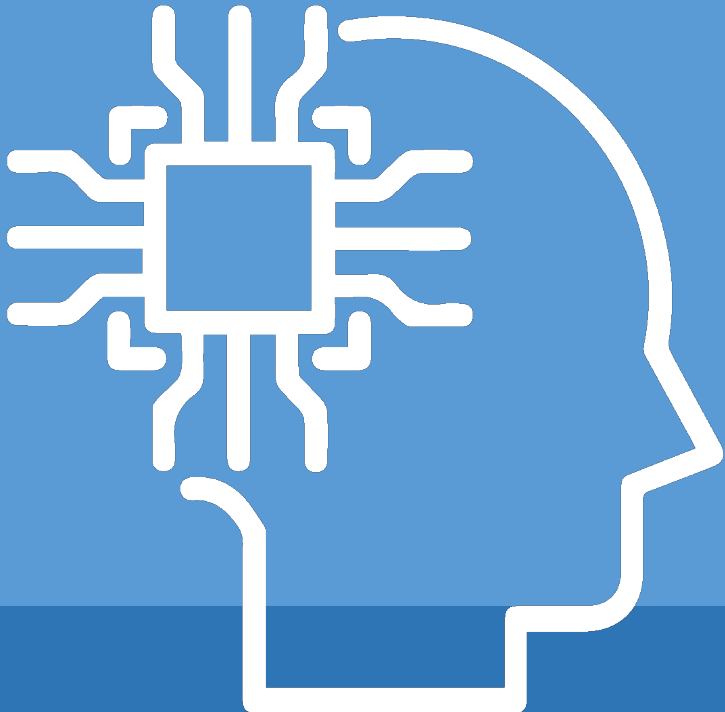


Student Course Survey (SCS)

May 14, 2024



Module:
**Perspectives
on HAI**



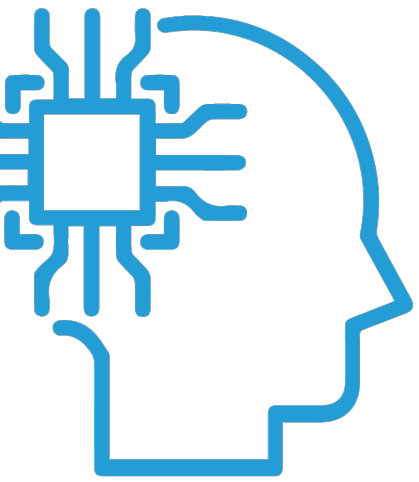
HISTORY OF AI

**Human Artificial Intelligence Interaction
(HAI)**

Iris Howley

Upcoming

- All upcoming deadlines are posted to the Glow page!
- Home (modules) has everything organized by Module
- Syllabus has everything organized by date



Glow Modules

Home Page (has announcements)

Fall 2022

Home

Announcements

Assignments

Quizzes

Discussions

Files

Course Media Gallery

Grades

Pages

People

Course Email

Syllabus

Modules

Collaborations

Libraries

Recent Announcements

Welcome! Modules 0 & 1 are posted!

Hi folks! Welcome to HAIL! Just a quick announcement to let you know that all of the lecture materials, t...

Posted on: Sep 11, 2022, 11:57 AM

Reply

Collapse All

▼ #0 Course Information (Getting Started in the Course) Complete All Items

Task List for Module #0 - starting 9/8

Attend Your Tues/Thurs Class

Lecture 1: Welcome!

Lecture00-1_CourseIntro.pdf

Start of Semester Survey

Read: How to Read & Understand a Scientific Paper

CSCI378syllabus-f22.pdf

Iris' Student Help Hours

Syllabus Quiz

▼ #1 Practical Machine Learning Complete All Items

Organized by Module (Topic)

Glow Syllabus

Organized by Deadline

23F-CSCI-378-01 - Human-AI Interaction > Syllabus

Presenter View

Jump to Today

Course Syllabus

CS378 Course Syllabus (a file) ↓

Below is a listing of Assignments due in this course. All Assignments will be posted to Glow, even if they are submitted through external means. Please be sure to click the Assignment link and read the details!

Please reach out to Iris if you have any concerns with the course. If I don't know about it, I can't help!

Course Summary:

Date	Details	Due
Tue Sep 13, 2022	Start of Semester Survey	due by 1am
	Syllabus Quiz	due by 9am
Sun Sep 18, 2022	Assignment 1 Check-in	due by 11:59pm
Sun Sep 25, 2022	Assignment 1: ML for Home Loans	due by 11:59pm
Fri Dec 9, 2022	Lead a Class Discussion on the Readings	due by 11:59pm
	Roll Call Attendance	

THIS LESSON

Historical AI

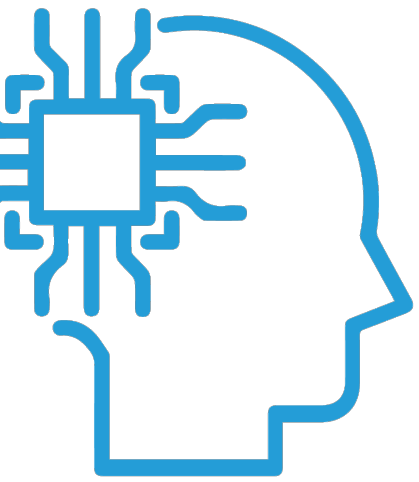
After this lesson, you should be able to:

1. Summarize the major hype/bust moments in AI
2. Describe key events forming the current wave of AI we see today

Topics in this Lecture

Or: Videos in this Playlist

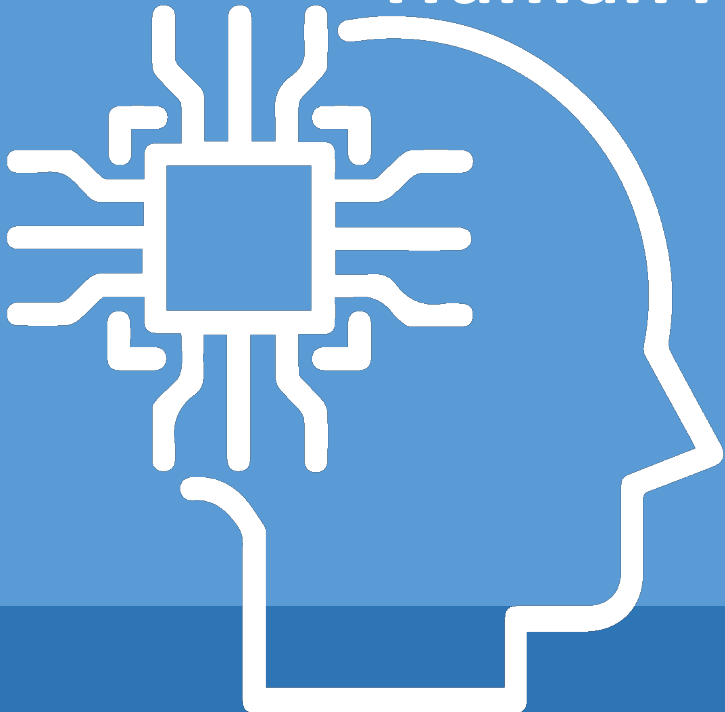
0. History of AI Introduction (this video!)
1. Ancient Artificial Intelligence
2. 1st AI Spring
3. 1st AI Winter
4. 2nd & 3rd AI Spring



Let's
GO!

Ancient AI

Human Artificial Intelligence Interaction (HAI)



Iris Howley

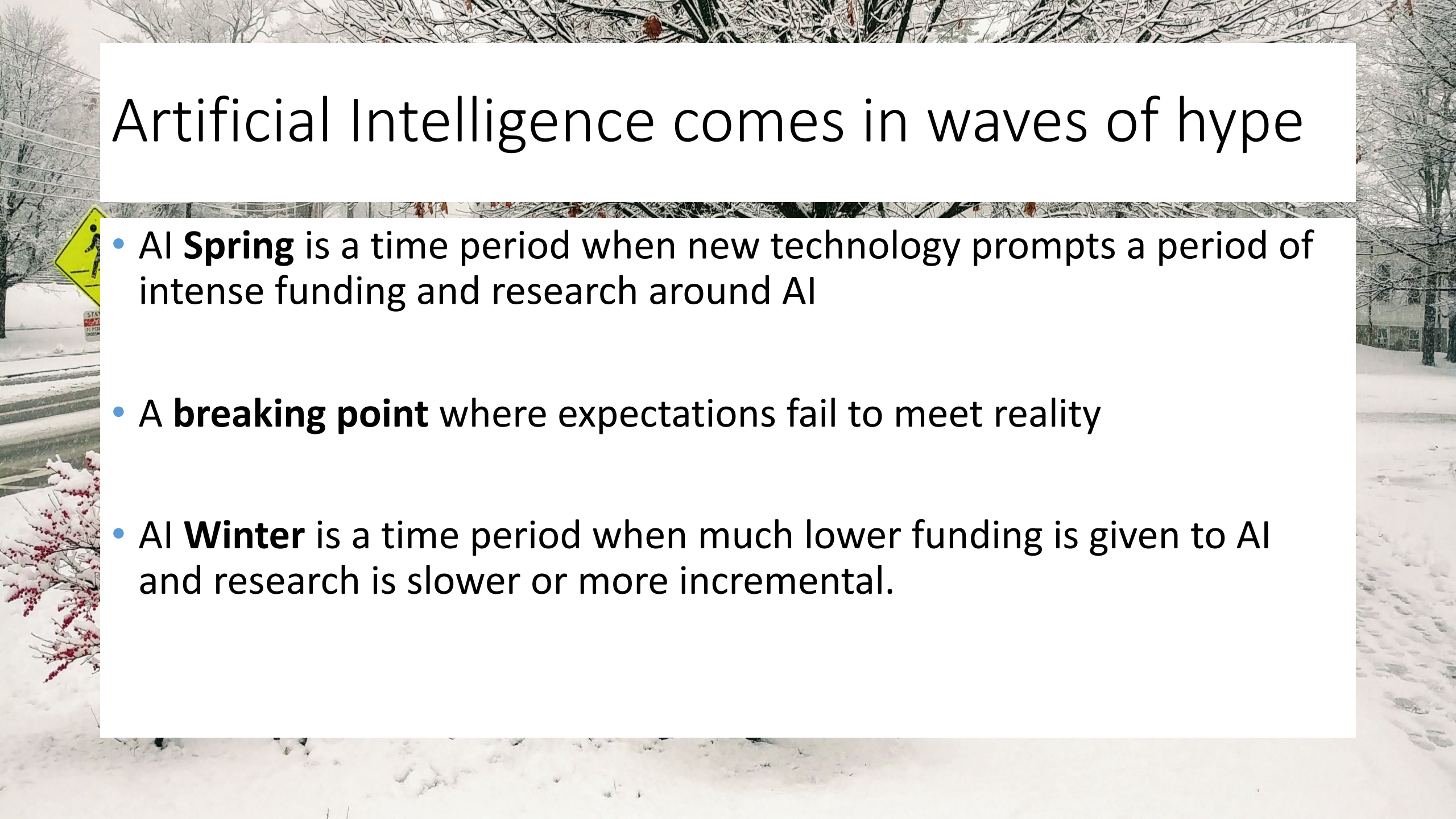
THIS LESSON

Historical AI.

(AI is an ooooooold concept)

Outline

- 0BC - 1956 Prehistory of Artificial Intelligence
- <Next Video!>
- 1956 - 1959 First Artificial Intelligence Spring
- <Next Video!>
- 1959 - 1974 First Artificial Intelligence Winter
- <Next Video!>
- 1980 - 1987 Second Artificial Intelligence Spring
- 2011 - Present Third Artificial Intelligence Spring

A background image of a snowy winter scene. In the foreground, there's a snow-covered ground with some red berries. In the middle ground, a yellow pedestrian crossing sign is visible on the left. The background shows snow-covered trees and a building in the distance.

Artificial Intelligence comes in waves of hype

- AI **Spring** is a time period when new technology prompts a period of intense funding and research around AI
- A **breaking point** where expectations fail to meet reality
- AI **Winter** is a time period when much lower funding is given to AI and research is slower or more incremental.

Ancient Automaton

"... by discovering the true nature of the gods, man has been able to reproduce it."

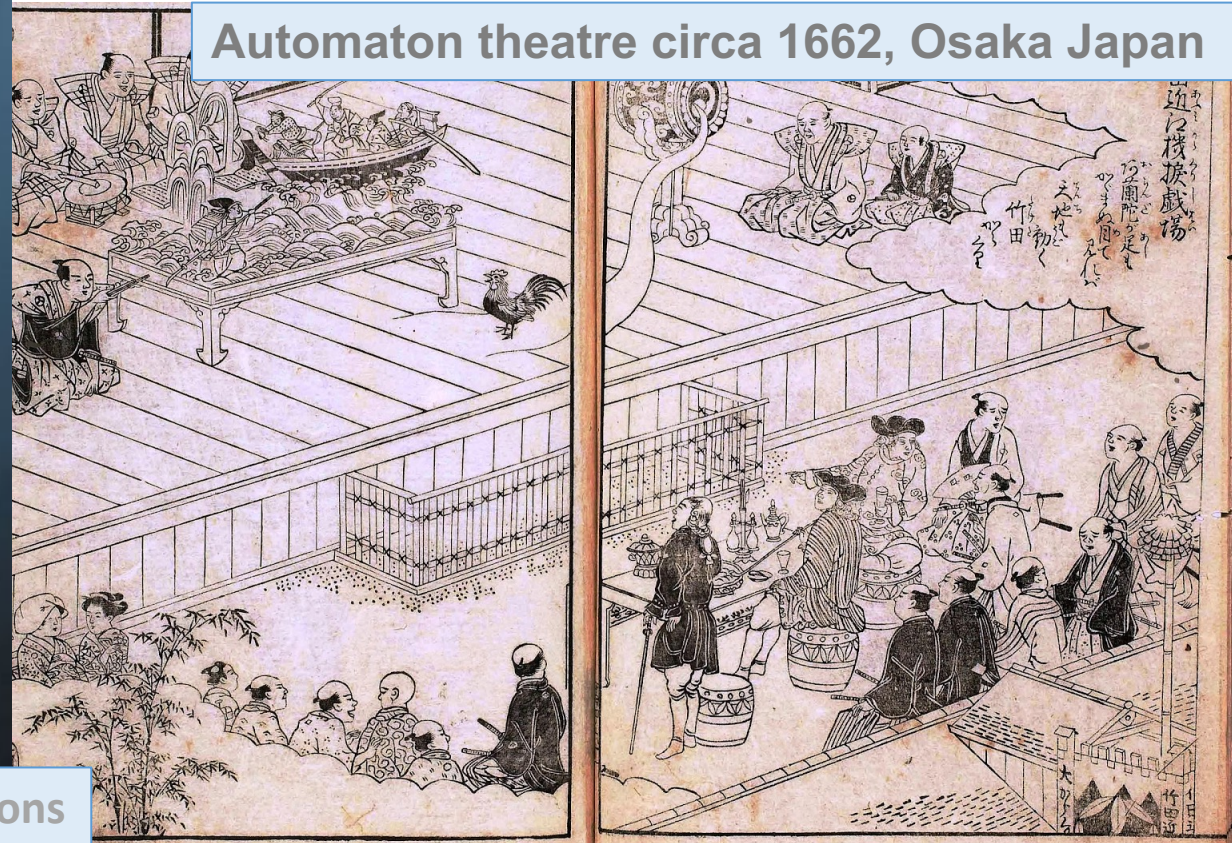
- maybe some guy named Hermes Trismegistus < 200 BC



Ancient Greek myth of Pygmalion



Hephaestus's metal automatons



Dreams of Robots 1860s - 1940s



“Metropolis” a 1927 Silent Film



“R.U.R” a 1920 play

**“...the time will come
when the machines will
hold the real supremacy
over the world.”**

- 1863 article by Samuel Butler

1942 Asimov's Laws

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.



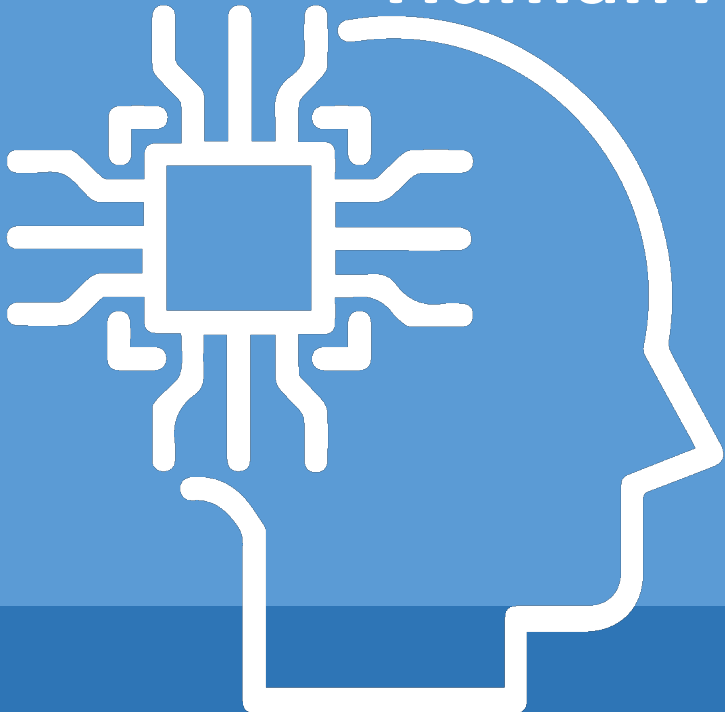
QUESTIONS?

Please contact me!

(more lecture follows...)

1st AI Spring

Human Artificial Intelligence Interaction (HAI)



Iris Howley

THIS LESSON

Up to the first AI winter.

(So we can avoid the mistakes of the past)

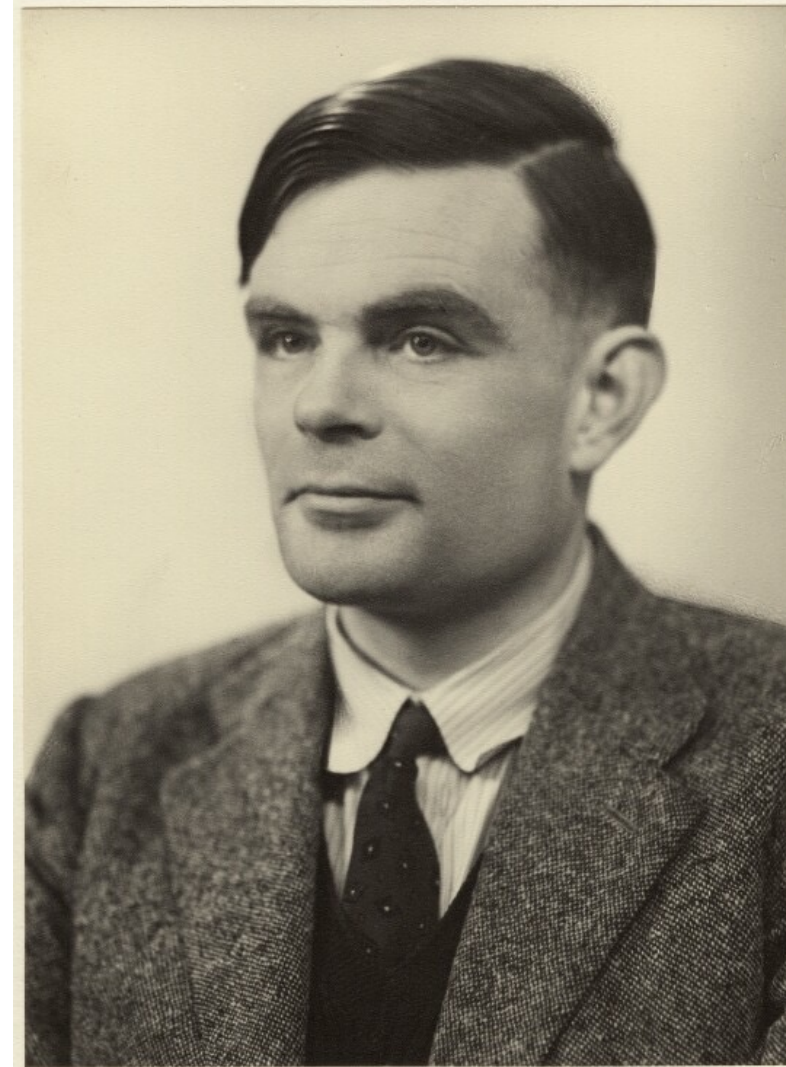
1950 - 1956 Beginnings of Computer Science

Lots of people: **What truly *is* human intelligence?**

Alan Turing: **How can we decide when a machine has achieved human-level intelligence?**

The politeness convention:

If a machine behaves as intelligently as a human being, then it is as intelligent as a human being.



Alan Turing

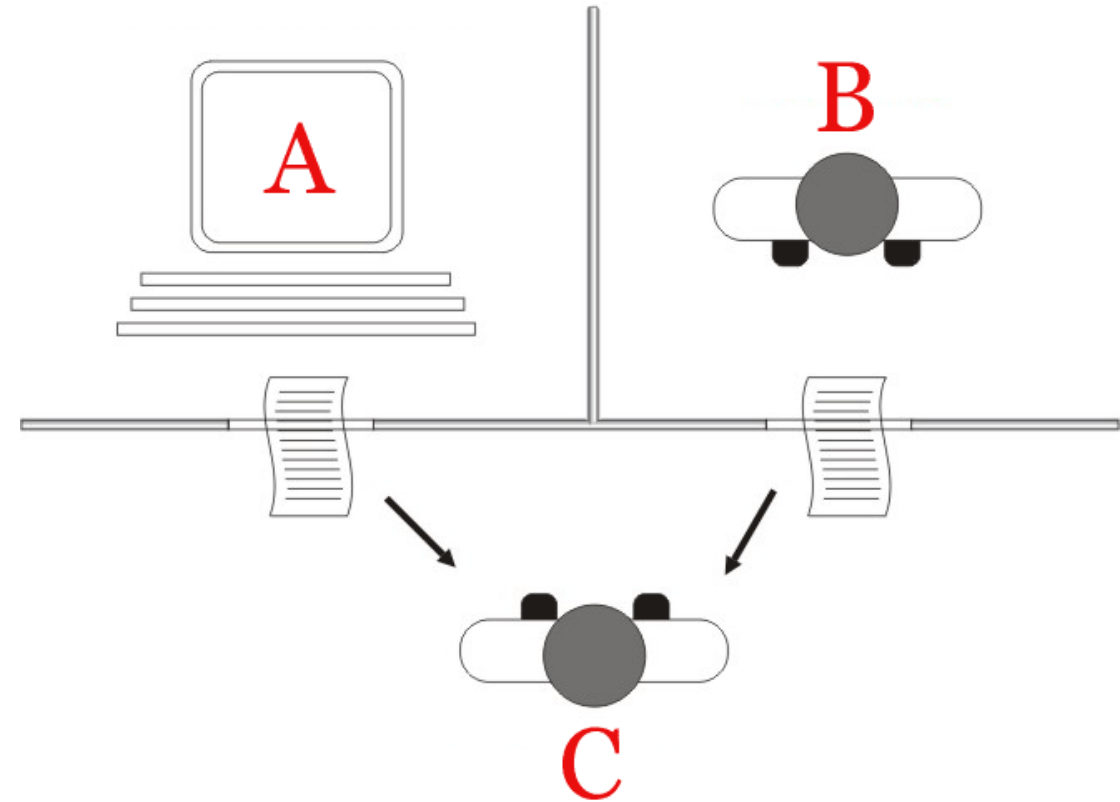
1950 - 1956 Beginnings of Computer Science

The Turing Test 1950

Player **C** uses natural language typing to hidden players **A** and **B**.

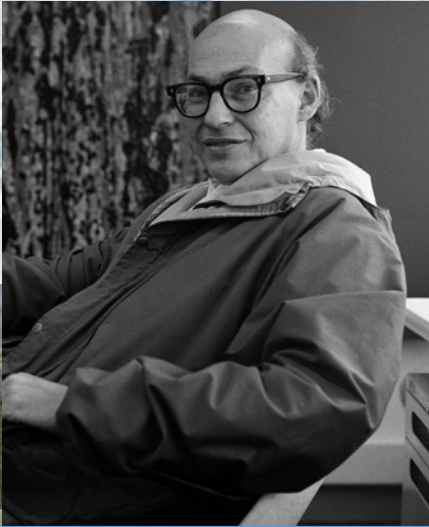
Player **C** attempts to uncover which player is human.

AI has been achieved if player **C** decides that the machine player is human.

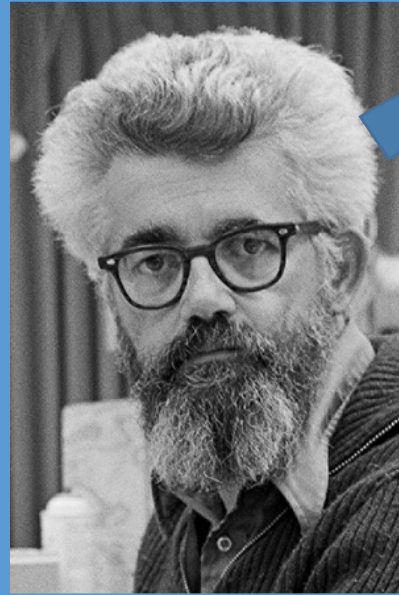


(Turing's death comes in 1954 at the age of 41)

AI Summer Camp: Researchers gather at Dartmouth in 1956 and Artificial Intelligence is born as a field.



Marvin Minsky, MIT



John McCarthy,
Dartmouth/Stanford



Herbert Simon,
CMU

Alan Newell,
CMU

1956 - 1971 We can teach innate knowledge through rules

Symbolism: formal logic systems can represent intelligent action

```
(RULE 5
  (IF (PCS-SCS HEAT TRANSFER INADEQUATE)
    (LOW FEEDWATER FLOW))
  (THEN (ACCIDENT IS LOSS OF FEEDWATER)))

(RULE 6
  (IF (SG INVENTORY INADEQUATE)
    (LOW FEEDWATER FLOW))
  (THEN (ACCIDENT IS LOSS OF FEEDWATER)))

(RULE 7
  (IF (PCS INTEGRITY CHALLENGED)
    (CONTAINMENT INTEGRITY CHALLENGED))
  (THEN (ACCIDENT IS LOCA)))

(RULE 8
  (IF (PCS INTEGRITY CHALLENGED)
    (SG LEVEL INCREASING))
  (THEN (ACCIDENT IS STEAM GENERATOR TUBE
    RUPTURE)))

(RULE 9
  (IF (SG INVENTORY INADEQUATE)
    (HIGH STEAM FLOW))
  THEN (ACCIDENT IS STEAM LINE BREAK)))
```

Figure 2. Event-oriented IF-THEN rules.

1956 - 1971 We can teach innate knowledge through rules

Symbolism: formal logic systems can represent intelligent action

- Newell & Simon's "General Problem Solver" can solve math proofs by searching a logic space
- Advances in natural language processing based on rules how words relate
- Advances in computer vision based on image transforms
- Advances in robotics based on rules and search in simplified settings

Don't make promises you can't keep

"Machines will be capable, within twenty years, of doing any work a man can do" - Herbert Simon, 1965

"In from three to eight years we will have a machine with the general intelligence of an average human being." - Minsky, 1970

1956 - 1971 We can teach computers to learn

Connectionism: computers should mimic how the brain works

- *Neurons* make thousands of links with other neurons, making *trillions* of possible connections in the brain
- An individual neuron will fire if specific input reaches a certain threshold of electricity, otherwise no
- Threshold for a neuron to fire = *activation weights* in a neural network

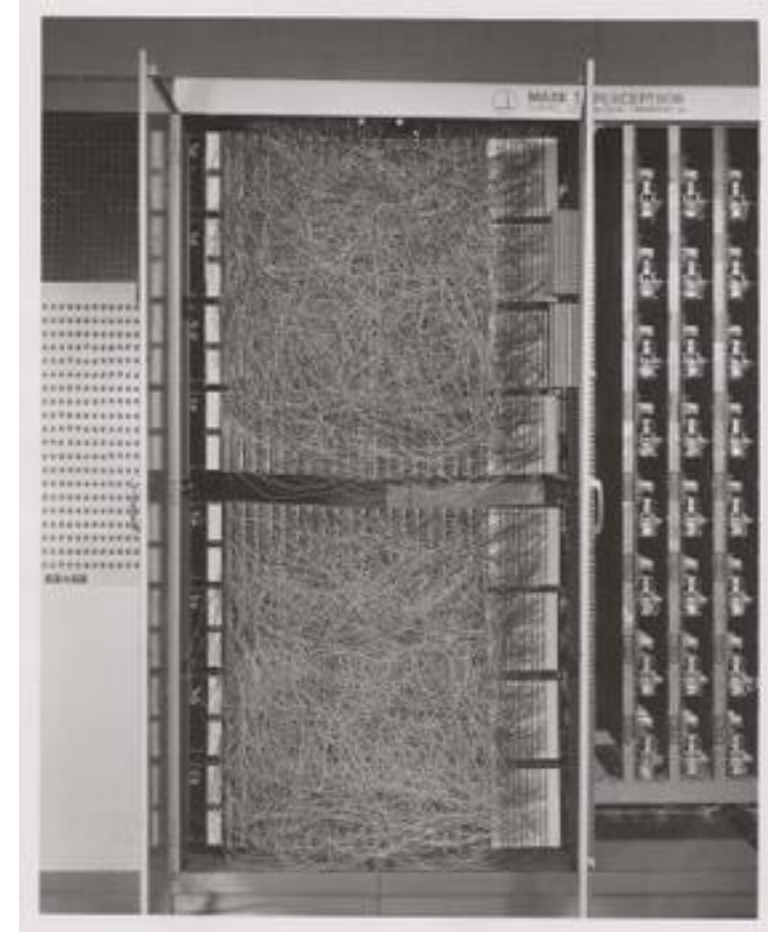
1956 - 1971 We can teach computers to learn

The Perceptron: designed by Frank Rosenblatt 1958.

"the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence."

- NYT 1958

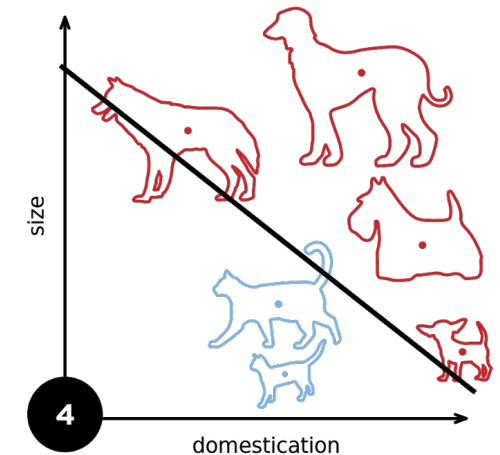
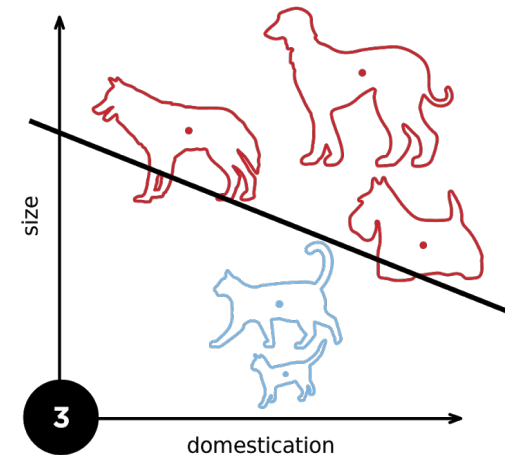
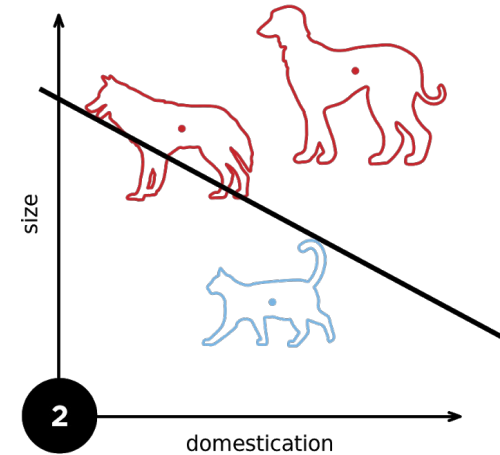
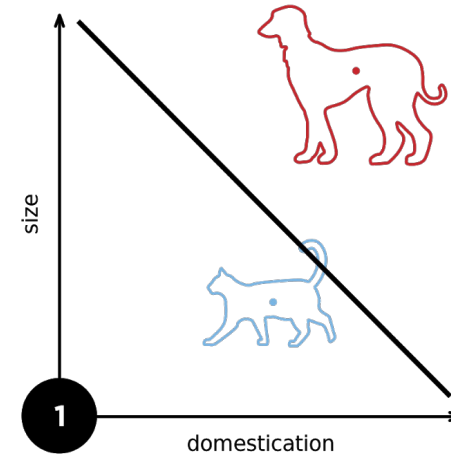
$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} + b > 0, \\ 0 & \text{otherwise} \end{cases}$$



1956 - 1971 We can teach computers to learn

The Perceptron

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} + b > 0, \\ 0 & \text{otherwise} \end{cases}$$



1959 Goodbye: the Perceptron destroyed

Minsky and Seymour Papert publish a book “Perceptrons”

- Burn piece of Perceptron approach in favor of rule approach
- Perceptron cannot handle XOR operator
- Shuts down funding for neural networks
- Rosenblatt soon dies, never to see neural nets re-vindicated



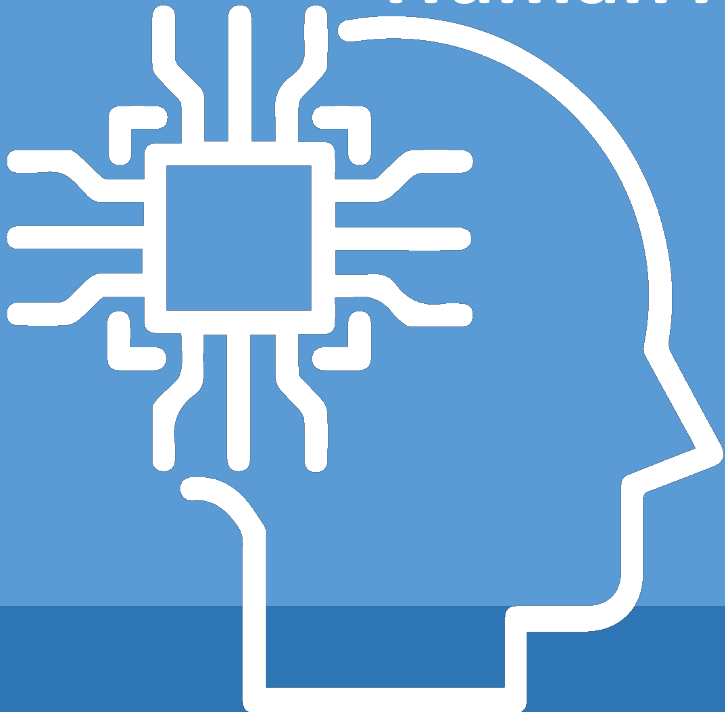
QUESTIONS?

Please contact me!

(more lecture follows...)

1st AI Winter

Human Artificial Intelligence Interaction (HAI)



Iris Howley

THIS LESSON

The first time AI lost its hype.

(So we can avoid the mistakes of the past)

First AI Winter

Funding is lost due to unmet promises

- Lighthill Report (1973) shuts down funding in UK
- ~1965 Dreyfus at MIT argues lots of human reasoning is *not based on logic rules*, involving instinct and unconscious reasoning
 - (No AI researcher will eat lunch with Dreyfus for the next decade)
- Sussman: “*using precise language to describe essentially imprecise concepts doesn't make them any more precise.*”

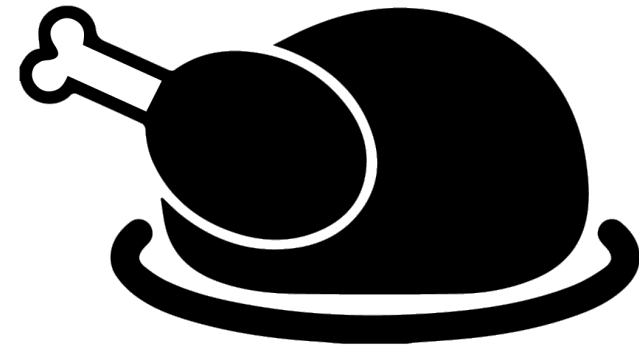
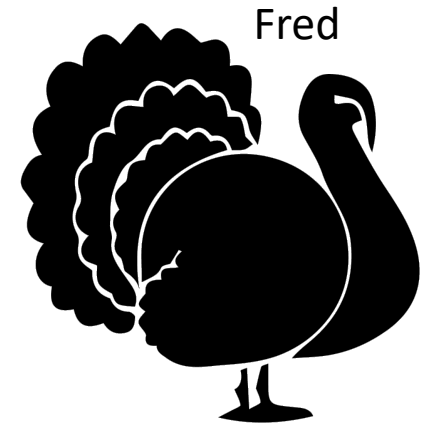
First AI Winter

Funding is lost due to unmet promises and **fundamental** limitations

- Computational power of the day limits most applications
- Lots of infighting
- Limitations with logic rules
 - All the possibilities
 - Common sense reasoning
 - Intractability & Combinatorial Explosion

Limitations with Logic rules: Yale shooting problem

1. Fred is alive and Alice has a gun
2. Alice loads the gun
3. Alice shoots at Fred
 - a. She missed
 - b. She shot in different direction
 - c. She hits in the wing
 - d. They're not at the same location
 - e. She thinks she shoots at Fred but it's a different Fred
 - f. It's a toy gun
 - g. The gun broke
 - h. Fred was revived, life saved
 - i. A duel
4. Fred is dead



If the first 3 events, is the 4th event true?

Limitations with Logic rules: Combinatorial Explosion

Example: robot motion planning



<https://www.youtube.com/watch?v=g0TaYhjpOfo>



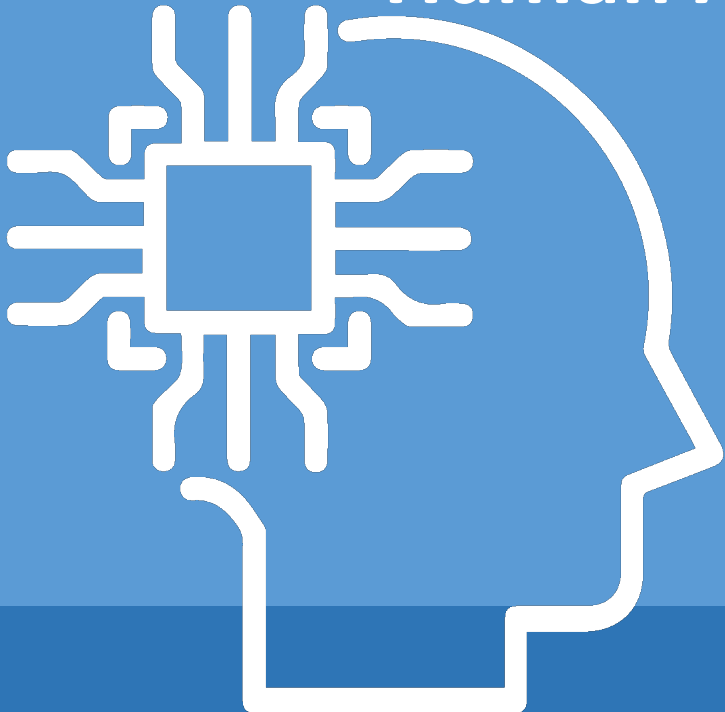
QUESTIONS?

Please contact me!

(more lecture follows...)

2nd & 3rd AI Spring

Human Artificial Intelligence Interaction (HAI)



Iris Howley

THIS LESSON

The more recent AI history.

(A better intuition about AI's general trajectory)

1980's from General Intelligence to Specific Applications

- Expert systems: while rules can't do everything they are useful for capturing what experts do
- 1989 Deep Thought at CMU can play human-level chess
- Again, hype gets too high and funding is lost

Can expert systems capture all expert-knowledge? What are the human-AI implications of expert systems?

1990s - 2000s AI undercover

- Many, many, many advances in computer science, probability theory, statistical learning, etc.
- **“Computer scientists and software engineers avoided the term artificial intelligence for fear of being viewed as wild-eyed dreamers.” - NYTimes 2005**
- Machine Learning as a field distances itself from AI

2010s - wait... now we can Machine Learn at scale

- Machine learning now immensely powerful
- Neural networks now practical, many earlier inventions in neural networks like back propagation in the 1980s are re-discovered
- Deep learning gives big performance leaps in almost all application areas overnight
- \$\$\$ massive funding is back

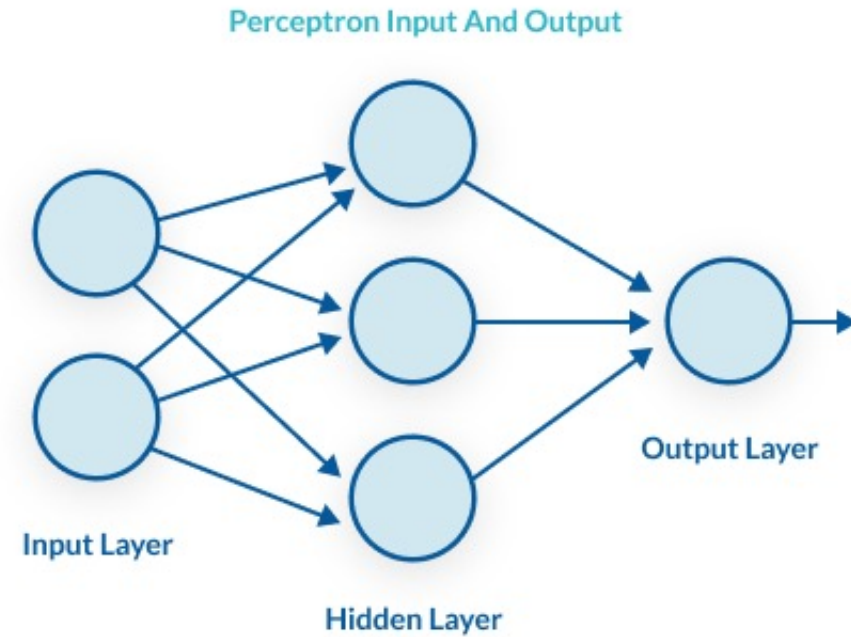
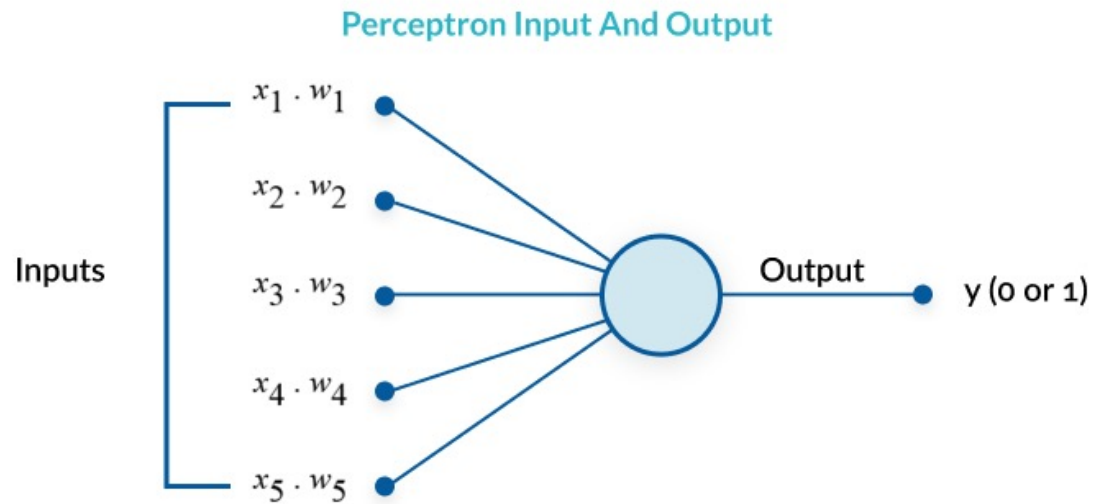
What factors in the 2010s enable a resurgence for AI?

2010s - wait... now we can Machine Learn at scale

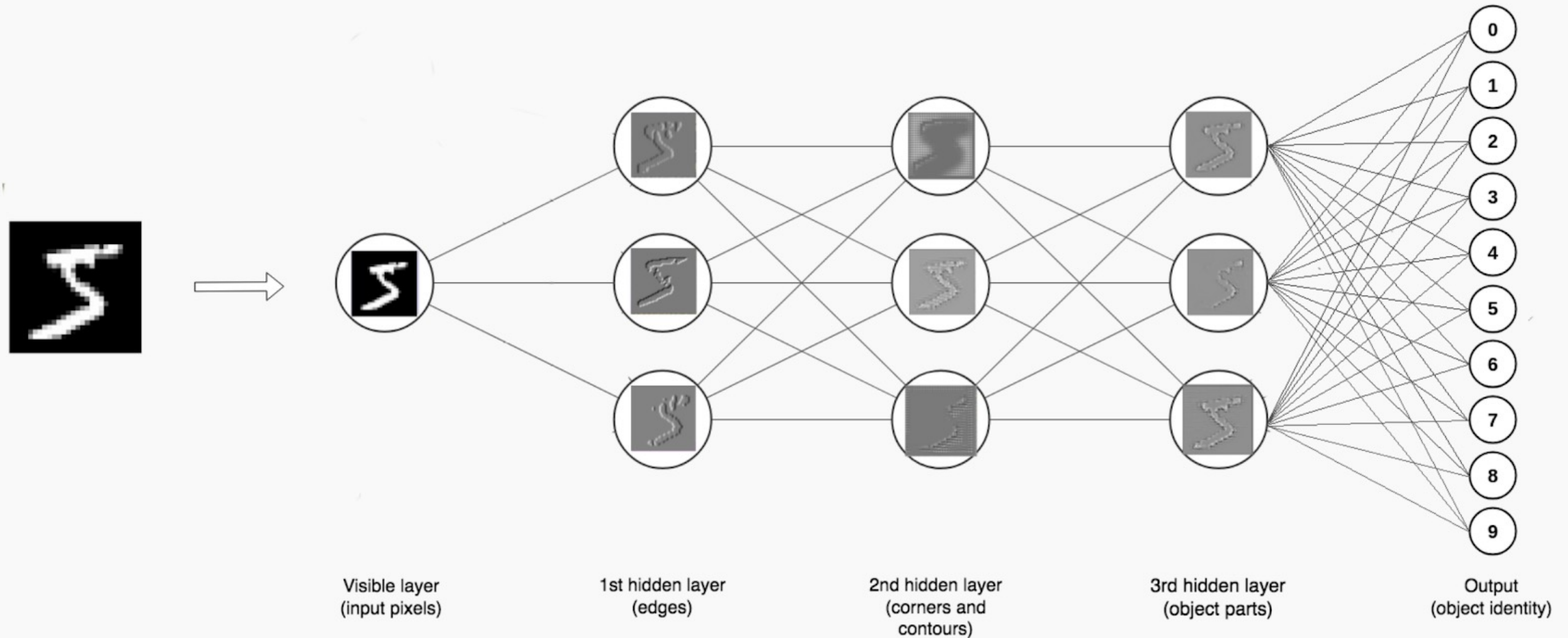
- Machine learning now immensely powerful
- Neural networks now practical, many earlier inventions in neural networks like back propagation in the 1980s are re-discovered
- Deep learning gives big performance leaps in almost all application areas overnight
- \$\$\$ massive funding is back

Why is human-AI interaction now so important?

What is a modern neural net vs. the 1970s Perceptron?



What is a modern neural net vs. the 1970s Perceptron?



Thought Exercise:

If machine learning approaches are so good, do we need rule-based approaches anymore?

- What advantages and disadvantages does each approach have?
- Do realistic applications need both? Or just one, based on the application?

Discussion Questions to Ponder

- Can expert systems capture all expert-knowledge? What are the human-AI implications of expert systems?
- What factors in the 2010s enabled a resurgence for AI?
- Why is human-AI interaction now so important?
- If machine learning approaches are so good, do we need rule-based approaches anymore?

Also, be prepared to discuss the readings!



Bring these with
you to Conference
Section!

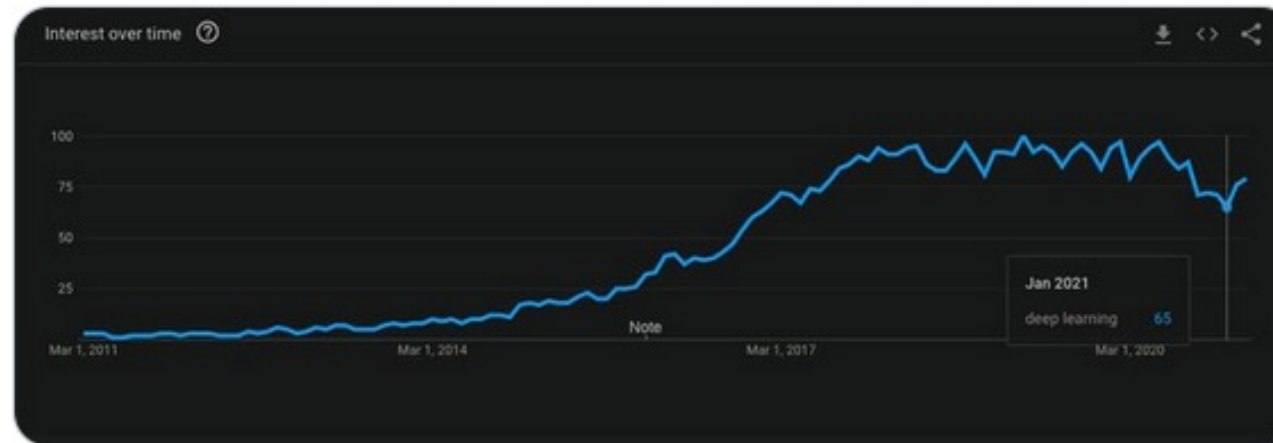
Now?



Andrew Trask
@iamtrask

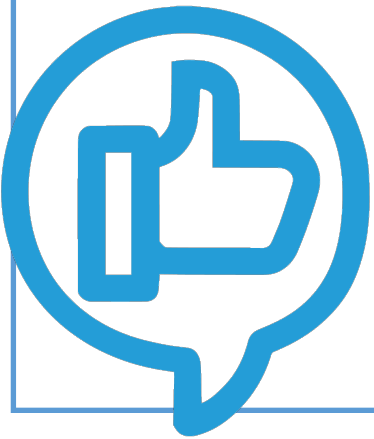


Google searches for "deep learning" are beginning to decline...



5:15 PM · Mar 5, 2021 · Twitter Web App

<https://twitter.com/iamtrask/status/1367961936109989900>

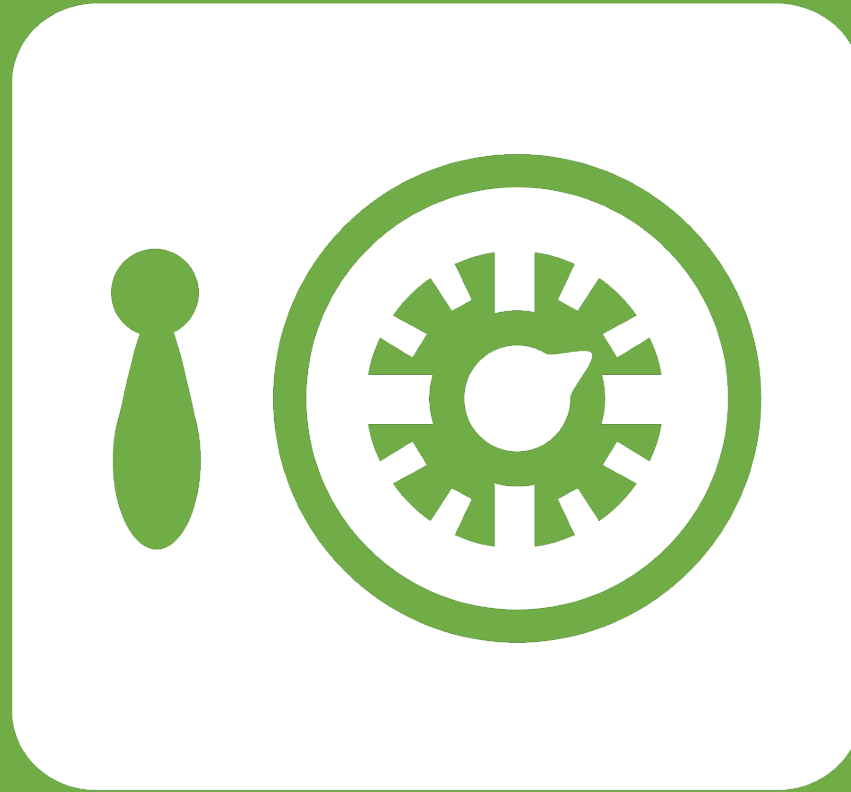




QUESTIONS?

Please contact me!

end.



Leftover Slides



REGISTER TO VOTE!

US Presidential election is November 3.
US Citizen and over 18 years old? REGISTER TO VOTE.

Turbovote has more info:

<https://turbovote.org/>