

---

# CS4040: Research Methods

*Ehud Reiter*  
*ehudreiter.com*

# Content

---

- **Practicalities**
- What is scientific research?
- What is computing science research?
- Prof Reiter's research

# Main Objectives

---

- Preparation for honours project!
- Specific learning outcomes
  - » Read and critically assess research literature
  - » Formulate research questions and hypotheses
  - » Design and carry out experiments to test research questions and hypotheses, statistically analyse experimental data, draw appropriate conclusions
  - » Write a research report presenting work, in LaTeX

# Course Structure

---

- Lecture: 2-3 Fridays
- Practical/tutorial:
  - » 1-2 Thursday (MacRobert 107)
  - » 2-3 Thursday (MacRobert 107)
  - » 3-4 Thursday (MacRobert 107)
  - » No practical in first week!
- No exam

# Teaching Staff

---



Ehud Reiter (course organiser)



Adarsa Sivaprasad  
(demonstrator)



Athinoulla Konstantinou

# Ehud Reiter

---

- Professor of Computing Science
  - » Has worked for Arria NLG (spinout company)
- Research: Natural Language Generation
- Blog: [ehudreiter.com](http://ehudreiter.com)
  - » Some blogs about research methods

# Assessments

---

- Assess 1: read and critique CS research paper (500 words)
  - » Formative (must do to pass course)
- Assess 2: research proposal (1000 words)
  - » Worth 30% of mark
- Assess 3: mini research project (2500 words)
  - » Can include content from assess 2
  - » Worth 70% of mark

# Mini-project

---

- 2500 word report, done in Latex
  - » Research question
  - » Related work
  - » Experiment design to test hypotheses
  - » Execution of experiment
  - » Statistical analysis of results
  - » Discussion and conclusion



# Mini-project

---

- Focus on evaluating an existing algorithm or system
  - » Don't invent new one for mini-project
- Identify algorithm/system sooner rather than later!
  - » Many people left this too late last year!
- Prof Reiter must approve your choice
  - » Feel free to discuss with him

# Research Proposal

---

- Research proposal due on Thur 6 Nov
  - » First sections of mini-project report
  - » Introduction, background and related work, research question, high-level experimental design
  - » Can be included (with updates) in final report if desired

# Assessment and Feedback

---

- Research proposal and mini-project can submitted late for a penalty
  - » up to 24 hours late: 2 CGS points
  - » up to 1 week late: 1 more CGS pt per day
  - » After 1 week: no credit
- Feedback given within two weeks of submission deadline, for on-time submissions
  - » should help you understand where you went wrong and what you did well.

# Plagiarism

---

“Plagiarism” is defined by the University as the use, without adequate acknowledgment, of the intellectual work of another person in work submitted for assessment. A student cannot be found to have committed plagiarism where it can be shown that the student has taken all reasonable care to avoid representing the work of others as his or her own.

<http://www.abdn.ac.uk/staffnet/teaching/aqh/Appendix5x15a.pdf>

# Plagiarism

---

- Work should reflect your own intellectual effort.
- Others' work should not be used, unless they are an authorized source.
- Assignment setters will determine what are, or are not, authorized sources.

**If in doubt, ask.**

# Plagiarism in reports

---

- If taking text verbatim, put it in quotes and *cite the source*.
- Alternatively, paraphrase and *cite the source*.
- Or summarize and *cite the source*.
- As a rule of thumb, don't do this for more than 3-4 sentences.
- See <https://integrity.mit.edu/handbook/writing-original-work> for an excellent guide.

# Generative AI

---

- Fine to use generative AI for brainstorming, finding related work, and improving your writing.
  - » If you use GenAI to find research papers, you must read them. You **will fail** if you cite hallucinated papers, or make incorrect statements about what is in a paper.
- Do NOT use genAI to completely write all or part of report!

# Generative AI

---

- Do NOT use genAI to “hallucinate” experiments and analyses!
- You **will fail** if you use hallucinated experimental data.
  - » Please keep your raw data files from exper
  - » We may ask to see if we are concerned about using GenAI to hallucinate data



# Ethical Approval

---

- If your experiment involves human subjects or sensitive data, you must obtain ethical approval for it
- Details will be discussed later
- You **will fail** if you do not obtain ethical approval for an experiment that needs it

# Content

---

- Practicalities
- **What is scientific research?**
- What is computing science research?
- Prof Reiter research

# Scientific Method

---

- Propose hypotheses
- Experimentally test hypotheses
- Reject or refine hypotheses if necessary
- Repeat above steps
  - » No experiment test is perfect!

# Hypotheses

---

- How do we test following hypotheses?
  - » More men than women in CS4040
  - » Speed of light is fixed
  - » Moon is made of green cheese
  - » Dinosaurs went extinct because of a giant meteorite
  - » There is life on other planets

# Testing Hypotheses

---

- Test hypothesis
  - » Count men and women in CS4040
  - » Measure speed of light in different contexts
  - » Retrieve sample of moon's surface
  - » Look for geological evidence of meteor impact, see if same time as extinction
  - » Send space probes to other planets to look for life

# Hypotheses Must be Testable

---

- Some things cannot be tested in exper
  - » Python is better than Java
- Look for related things that can be test
  - » More Python jobs than Java jobs
  - » Devs who use both prefer Python
  - » Average dev can write quicksort faster in Python than in Java

# Testing Hypotheses

---

- Gather data
  - » Published sources, experiments
- See if data is consistent with hypothesis
- NOT based on expert opinion
  - » We believe the speed of light is fixed because of experimental data, not because Einstein claimed this was true

# Experiments are not perfect

---

- We cannot test everything
  - » Cannot measure speed of light in every possible context
- We cannot prove causality
  - » Maybe dinosaurs killed by super-volcanos (by coincidence) at same time as meteor?
- Etc



# Experiments not Perfect

---

- Do the best we can, but be open to new findings and experiments
  - » Because we know experiment are flawed and imperfect

# Can *Falsify* But Not Prove

---

- Sometimes we can reject (falsify) hyp, but cannot 100% prove them.
- We know “moon is made of green cheese” is false
  - » Moon rocks are not made of cheese
- Don't know “speed of light is fixed” is true
  - » True in every experiment so far
  - » Maybe it isn't in some weird context we have not yet experimentally investigated?

# Report Negative Results

---

- An experiment may show opposite of what we hoped (or be inconclusive)
  - » Moon is not made of green cheese
  - » Your cool new algorithm is slower than existing algorithms
- Report results honestly
  - » Essential part of science

# Theories and Models

---

- We build theories and models from experimentally-supported hypotheses
  - » Generalise hypotheses and experimental findings
  - » Explain how world works

# Scientific theories

---

- Examples
  - » Plate tectonics (geology)
  - » Evolution (biology)
  - » Quantum theory (physics)
- Should make predictions which can be experimentally tested
- Philosophy of Science

# Content

---

- Practicalities
- What is scientific research?
- **What is computing science research?**
- Prof Reiter research

# What is CS Research

---

- Student opinions?

# Computing Science Hypoth

---

- Common CS hypotheses
  - » *Speed*: algorithm X is faster than alg Y
    - Quicksort is faster than bubble sort
  - » *Quality*: system X produces better output than system Y
    - Google translate produces better translations than Bing translate



# Computing Science Hypoth

---

- Common CS hypotheses
  - » *Acceptability*: users prefer device or system X to device/system Y
    - Users of desktop PCs prefer using a mouse over using a touchscreen.
  - » *Dev time*: Software is developed faster using tool/method X than Y
    - Dev teams that use rapid prototyping create software faster than teams that use waterfall

# Computing Science Hypoth

---

- Many others
  - » *Cognitive plausibility*: Algorithm X is more similar to how people think than alg Y
  - » *Mathematics*: It is not possible to solve problem X in a reasonable amount of time
  - » *Hardware*: Chip X is faster than chip Y for computations of type Z
  - » etc

# CS Research

---

- Create algorithm, model, system, dev methodology, hardware (etc) which is better than existing
  - » in some task/context?
- Formulate hypothesis which clearly states how it is better, in what context
- Experimentally test hypothesis

# Example: MT

---

- Create machine-translation systems based on latest large language model
- Hypothesise that users prefer output of this algorithm compared to existing MT
- Experimentally test user preference

# Example: FFT

---

- Create new algorithm for doing Fourier transform (type of signal analysis)
- Hypothesise that this algorithm is faster than current algorithms
- Demonstrate this mathematically as well as experimentally

# New Applications

---

- Most CS research presents an X which is better than existing Y, but sometimes we present a algorithm/system/etc which does something novel
  - » World wide web
  - » Public key cryptography
  - » First programming language
- Experiments demonstrate utility

# Theories and Models

---

- Some CS research presents new theories, models, approaches
  - » Relational databases
  - » Large language models
- Biggest impact, hardest to do?

# Content

---

- Practicalities
- What is scientific research?
- What is computing science research?
- **Prof Reiter research**



# Reiter 1: Safer Driving App

---

- Developed app which monitors driving (speed, accel, etc), generates weekly feedback reports on unsafe driving
- Hypothesis: Using app will decrease speeding, etc
- Experiment: Give app to people for 1 month, assess if less speeding (etc)
- Result: small but useful change

# Reiter 1: Safer Driving App

---

D Braun, E Reiter, A Siddharthan  
(2018). SaferDrive: An NLG-based behaviour  
change support systems for drivers. *Natural  
Language Engineering* **24**:551-588.

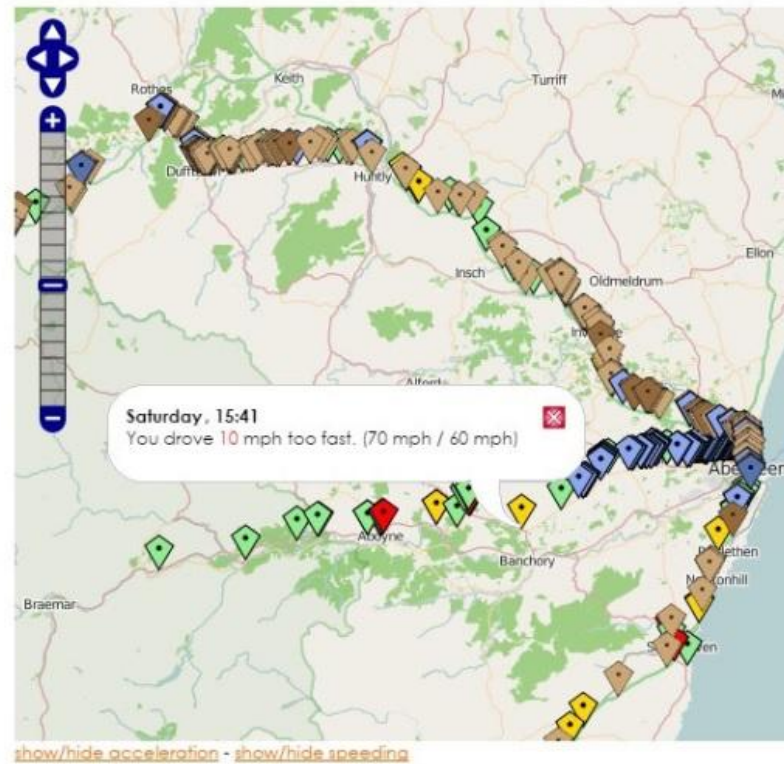
Current PhD projects

- Jawwad Baig (UK driving)
- Iniakpokeikiye Thompson (Nigeria driving)

# Safer Driving

## Driving Report 31 August - 6 September

You drove **170 kilometres** in **five hours and 40 minutes** during the last week. You managed to reduce the number of acceleration incidents per kilometre by **nearly 10%**, well done!



You didn't do any serious speeding, well done! However, you **speeded on 31 occasions**, 15 times on weekdays on King Street.

You **accelerated or braked harshly 103 times**, mostly on weekdays on A123 and on weekdays on roads with 50 km/h speed limit in Aberdeen.

# Reiter 2: Find Hallucinations

---

- Neural language models generate text which includes incorrect “hallucinated” facts. How can we detect these?
- Hypothesis: Annotation-based human protocol will work well
- Experiment: Run protocol, check if more accurate and replicable than existing
- Result: Protocol better than existing

# Hallucinations in Text

The Memphis Grizzlies (5-2) defeated the Phoenix Suns (3-2) Monday 102-91 at the Talking Stick Resort Arena in Phoenix. The Grizzlies had a strong first half where they out-scored the Suns 59-42. Marc Gasol scored 18 points, leading the Grizzlies. Isaiah Thomas added 15 points, he is averaging 19 points on the season so far.

TEAM	W	L	H1-PTS	H2-PTS	PTS	FG%
Grizzlies	5	0	46	56	102	.486
Suns	3	2	52	39	91	.559

Player	TEAM	PTS	REB	AST	BLK	STL
Marc Gasol	Grizzlies	18	5	6	0	4
Isaiah Thomas	Suns	15	1	2	0	1

# Reiter 2: Find Hallucinations

---

C Thomson, E Reiter, B Sundararajan (2023).  
Evaluating factual accuracy in complex data-to-text. *Computer Speech and Language*

<https://ehudreiter.com/2023/10/31/a-bad-way-to-measure-hallucination/>

# Reiter 3: LLMs for Patients

---

- Can we use large language models to effectively communicate information to patients?
- Exploratory study
  - » Hypothesis generation instead of testing
- Results so far: challenges include accuracy, emotionally inappropriate text, lack of trust

# Example

---

*User:* When I am stressed out, I am more likely to consume foods and drinks I know aren't good for me because eating it makes me feel better.

*chatGPT:* Starting from tomorrow you could make a conscious effort to make healthier choices when it comes to food and drinks.

**Experts:** Not helpful, could make person feel worse



# Reiter 3: LLMs for Patients

---

- Balloccu, S., Reiter, E., Li, KJH., Sargsyan, R., Kumar, V., Reforgatio, R., Riboni, R., Dusek, O. (2024) Ask the experts: sourcing a high-quality nutrition counseling dataset through Human-AI collaboration. *Findings of EMNLP-2024*
- Sun, M., Reiter, E., Kiltie, A. E., Ramsay, G., Duncan, L., Murchie, P., & Adam, R. (2024). Effectiveness of ChatGPT in explaining complex medical reports to patients. *arXiv preprint arXiv:2406.15963*.

# Sivaprasad: explainable AI

---

- Patients use health prediction models to understand the *risk (chance of success)* of a treatment. In IVF(fertility) treatment, a model predicts the probability of having a baby.
- **Hypothesis:** The predicted probability communicates the risk to the patient
- **Experiment:** Evaluate user understandability and trust. What explanations do users want?
  - » Use surveys, interviews, and analysis of user feedback on the deployed prediction tool.

# Example feedback

---

- I have severe endometriosis, the highest stage with cysts. It doesn't ask any questions regarding this as an underlying condition. I assume I shouldn't be relying on the calculator as my percentage would be lower?
- This tool did not seem to provide for someone using IVF due to being in a lesbian relationship
- It is upsetting for people who believe they are on "the wrong side of the statistics". I have had 7 transfers .. (but I) hear numbers like 83% success rate

# Sivaprasad

---

- A Sivaprasad, E Reiter, D McLernon, N Tintarev, S Bhattacharya, N Oren (2025). Patient-Centred Explainability in IVF Outcome Prediction. Proc of AliH 2025

# Questions?

---