# Presenting your research

Christopher Potts

Stanford Linguistics

CS224u: Natural language understanding

# Your papers

# Final paper details

https://web.stanford.edu/class/cs224u/projects.html#paper

https://github.com/cgpotts/cs224u/blob/master/projects.md

https://web.stanford.edu/class/cs224u/restricted/past-final-projects/

# Your projects (a reminder)

1. We will never evaluate a project based on how “good” the results are.
   - Publication venues do this, because they have additional constraints on space that lead them to favor positive evidence for new developments over negative results.
   - In CS224u, we are not subject to this constraint, so we can do the right and good thing of valuing positive results, negative results, and everything in between.
2. We will evaluate your project on:
   - The appropriateness of the metrics.
   - The strength of the methods.
   - The extent to which the paper is open and clear-sighted about the limits of its findings.

# Known project limitations

Imagine that your reader is a well-intentioned NLP practitioner who is seeking to make use of your data, models, or findings as part of a separate scholarly project, deployed system, or some other kind of real-world intervention. What should such a person know about your work?

## Examples of things you could include

- Benefits and risks
- Costs to your participants, society, the planet
- Responsible use of your data, models, findings

## Other resources

- Datasheets: Gebru et al. 2018
- Model cards: Mitchell et al. 2019
- Survey of NeurIPS impact statements: Nanayakkara et al. 2021

# Authorship statement

1. Explain how the individual authors contributed to the project.
2. You are free to include whatever information you deem important to convey.
3. Model: `http://blog.pnas.org/iforc.pdf` (p. xiii).
4. Rationale: we think this is an important aspect of scholarship in general.
5. Only in extreme cases, and after discussion with the team, would we consider giving separate grades to team members based on this statement.

# Policy on multiple submissions

`http://web.stanford.edu/class/cs224u/requirements.html#multiple`

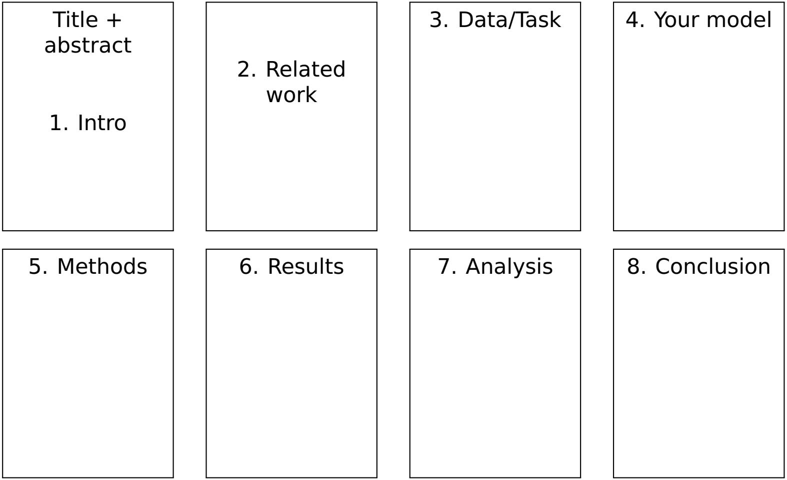
Notes:

1. Mirrors the policy on multiple submission to conferences.
2. Designed to ensure that your project is a substantial new effort.
3. Yes, this *does* mean that you can’t merely submit an incremental advancement over another project you did.
4. Other courses might have different policies, but that fact alone will not lead us to change our policy.
5. If any of these policies seem relevant to your work, start the discussion with your mentor as early as possible.

# Writing NLP papers

# The outline of a typical NLP paper

Four or eight two-column pages not including references.
Here are the typical components (section lengths will vary):



Title + abstract

1. Intro

2. Related work

3. Data/Task

4. Your model

5. Methods

6. Results

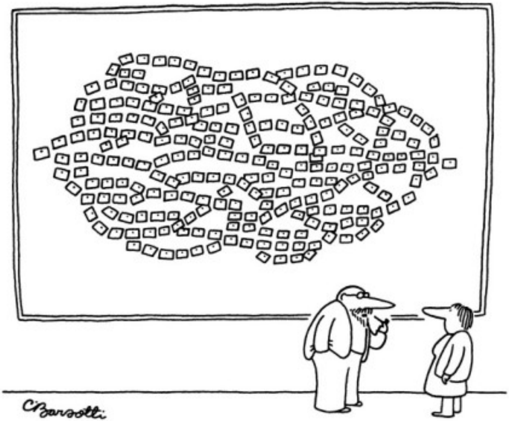
7. Analysis

8. Conclusion

# Additional notes

1. Intro: Tell the full story of your paper at a high-level.
2. Related work: Contextualize your work and provide insights into major relevant themes of the literature as a whole. Use each paper (or theme) as a chance to articulate what is special about your paper.
3. Data: Likely to be very detailed if the datasets are new or unfamiliar to the community, or if familiar datasets are being used in new ways.
4. Your model: Flesh out your own approach, perhaps amplifying themes from the ‘Prior lit’ section.
5. Methods: The experimental approach, including descriptions of metrics, baseline models, etc. Details about hyperparameters, optimization choices, etc., are probably best given in appendices, unless they are central to the arguments.
6. Results: A no-nonsense report of what happened.
7. Analysis: Discussion of what the results mean, what they don’t mean, where they can be improved, etc. These sections vary a lot depending on the nature of the paper.
8. (For papers reporting on experiments with multiple datasets, it can be good to repeats Methods/Results/Analysis in separate (sub)sections for each dataset.)
9. Conclusion: Quickly summarize what the paper did, and then chart out possible future directions that anyone might pursue.

# General advice on scientific writing



*"It's plotted out. I just have to write it."*

# Stuart Shieber: the ‘rational reconstruction’

- Continental style: “in which one states the solution with as little introduction or motivation as possible, sometimes not even saying what the problem was” [. . . ] “Readers will have no clue as to whether you are right or not without incredible efforts in close reading of the paper, but at least they’ll think you’re a genius.”
- Historical style: “a whole history of false starts, wrong attempts, near misses, redefinitions of the problem.” [. . . ] “This is much better, because a careful reader can probably follow the line of reasoning that the author went through, and use this as motivation. But the reader will probably think you are a bit addle-headed.”
- Rational reconstruction: “You don’t present the actual history that you went through, but rather an idealized history that perfectly motivates each step in the solution.” [. . . ] “The goal in pursuing the rational reconstruction style is not to convince the reader that you are brilliant (or addle-headed for that matter) but that **your solution is trivial**. It takes a certain strength of character to take that as one’s goal.”

[link]

# David Goss’s hints on mathematical style

“Have mercy on the reader.”

# Cormac McCarthy

Lots of good advice. The piece I want to highlight:

> *Decide on your paper's theme and two or three points you want every reader to remember. This theme and these points form the single thread that runs through your piece. The words, sentences, paragraphs and sections are the needlework that holds it together. If something isn't needed to help the reader to understand the main theme, omit it.*

This strategy will not only result in a better paper, but it will also be an easier paper for you to write, since the themes you choose will determine what to include/exclude and resolve a lot of low-level questions about the narrative.

https://www.nature.com/articles/d41586-019-02918-5

# Honesty

Patrick Blackburn's fundamental insight:

Where do good talks come from?

Honesty.

"A good talk should never stray far from simple, honest communication."

https://web.stanford.edu/class/cs224u/readings/blackburn2001.pdf

# A look at two really well-written papers

### Deep contextualized word representations

**Matthew E. Peters**[†], **Mark Neumann**[†], **Mohit Iyyer**[†], **Matt Gardner**[†],
{matthewp,markn,mohiti,mattg}@allenai.org

**Christopher Clark**[*], **Kenton Lee**[*], **Luke Zettlemoyer**[†*]
{csquared,kentonl,lsz}@cs.washington.edu

[†]Allen Institute for Artificial Intelligence
[*]Paul G. Allen School of Computer Science & Engineering, University of Washington

#### Abstract

We introduce a new type of *deep contextualized* word representation that models both (1) complex characteristics of word use (e.g., syntax and semantics), and (2) how these uses vary across linguistic contexts (i.e., to model polysemy). Our word vectors are learned functions of the internal states of a deep bidirectional language model (biLM), which is pretrained on a large text corpus. We show that these representations can be easily added to existing models and significantly improve the state of the art across six challenging NLP problems, including question answering, textual entailment and sentiment analysis. We also present an analysis showing that exposing the deep internals of the pre-trained network is crucial, allowing downstream models to mix different types of semi-supervision signals.

#### 1 Introduction

Pre-trained word representations (Mikolov et al., 2013; Pennington et al., 2014) are a key component in many neural language understanding models. However, learning high quality representations can be challenging. They should ideally model both (1) complex characteristics of word use (e.g., syntax and semantics), and (2) how these uses vary across linguistic contexts (i.e., to model polysemy). In this paper, we introduce a new type of *deep contextualized* word representation that directly addresses both challenges, can be easily integrated into existing models, and significantly improves the state of the art in every considered case across a range of challenging language understanding problems.

Our representations differ from traditional word type embeddings in that each token is assigned a representation that is a function of the entire input sentence. We use vectors derived from a bidirectional LSTM that is trained with a coupled language model (LM) objective on a large text corpus. For this reason, we call them ELMo (Embeddings from Language Models) representations. Unlike previous approaches for learning contextualized word vectors (Peters et al., 2017; McCann et al., 2017), ELMo representations are deep, in the sense that they are a function of all of the internal layers of the biLM. More specifically, we learn a linear combination of the vectors stacked above each input word for each end task, which markedly improves performance over just using the top LSTM layer.

Combining the internal states in this manner allows for very rich word representations. Using intrinsic evaluations, we show that the higher-level LSTM states capture context-dependent aspects of word meaning (e.g., they can be used without modification to perform well on supervised word sense disambiguation tasks) while lower-level states model aspects of syntax (e.g., they can be used to do part-of-speech tagging). Simultaneously exposing all of these signals is highly beneficial, allowing the learned models select the types of semi-supervision that are most useful for each end task.

Extensive experiments demonstrate that ELMo representations work extremely well in practice. We first show that they can be easily added to existing models for six diverse and challenging language understanding problems, including textual entailment, question answering and sentiment analysis. The addition of ELMo representations alone significantly improves the state of the art in every case, including up to 20% relative error reductions. For tasks where direct comparisons are possible, ELMo outperforms CoVe (McCann et al., 2017), which computes contextualized representations using a neural machine translation encoder. Finally, an analysis of both ELMo and CoVe reveals that deep representations outperform

2227

*Proceedings of NAACL-HLT 2018*, pages 2227–2237
New Orleans, Louisiana, June 1 - 6, 2018. ©2018 Association for Computational Linguistics

[link]

### GloVe: Global Vectors for Word Representation

**Jeffrey Pennington, Richard Socher, Christopher D. Manning**
Computer Science Department, Stanford University, Stanford, CA 94305
jpennin@stanford.edu, richard@socher.org, manning@stanford.edu

#### Abstract

Recent methods for learning vector space representations of words have succeeded in capturing fine-grained semantic and syntactic regularities using vector arithmetic, but the origin of these regularities has remained opaque. We analyze and make explicit the model properties needed for such regularities to emerge in word vectors. The result is a new global log-bilinear regression model that combines the advantages of the two major model families in the literature: global matrix factorization and local context window methods. Our model efficiently leverages statistical information by training only on the nonzero elements in a word-word co-occurrence matrix, rather than on the entire sparse matrix or on individual context windows in a large corpus. The model produces a vector space with meaningful substructure, as evidenced by its performance of 75% on a recent word analogy task. It also outperforms related models on similarity tasks and named entity recognition.

#### 1 Introduction

Semantic vector space models of language represent each word with a real-valued vector. These vectors can be used as features in a variety of applications, such as information retrieval (Manning et al., 2008), document classification (Sebastiani, 2002), question answering (Tellex et al., 2003), named entity recognition (Turian et al., 2010), and parsing (Socher et al., 2013).

Most word vector methods rely on the distance or angle between pairs of word vectors as the primary method for evaluating the intrinsic quality of such a set of word representations. Recently, Mikolov et al. (2013c) introduced a new evaluation scheme based on word analogies that probes the finer structure of the word vector space by examining not the scalar distance between word vectors, but rather their various dimensions of difference. For example, the analogy "king is to queen as man is to woman" should be encoded in the vector space by the vector equation *king* − *queen* = *man* − *woman*. This evaluation scheme favors models that produce dimensions of meaning, thereby capturing the multi-clustering idea of distributed representations (Bengio, 2009).

The two main model families for learning word vectors are: 1) global matrix factorization methods, such as latent semantic analysis (LSA) (Deerwester et al., 1990) and 2) local context window methods, such as the skip-gram model of Mikolov et al. (2013c). Currently, both families suffer significant drawbacks. While methods like LSA efficiently leverage statistical information, they do relatively poorly on the word analogy task, indicating a sub-optimal vector space structure. Methods like skip-gram may do better on the analogy task, but they poorly utilize the statistics of the corpus since they train on separate local context windows instead of on global co-occurrence counts.

In this work, we analyze the model properties necessary to produce linear directions of meaning and argue that global log-bilinear regression models are appropriate for doing so. We propose a specific weighted least squares model that trains on global word-word co-occurrence counts and thus makes efficient use of statistics. The model produces a word vector space with meaningful substructure, as evidenced by its state-of-the-art performance of 75% accuracy on the word analogy dataset. We also demonstrate that our methods outperform other current methods on several word similarity tasks, and also on a common named entity recognition (NER) benchmark.

We provide the source code for the model as well as trained word vectors at http://nlp.stanford.edu/projects/glove/.

1532

*Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543,
October 25-29, 2014, Doha, Qatar. ©2014 Association for Computational Linguistics

[link]

# NLP conference submissions

# The ACL anonymity period

1. The ACL conferences have adopted a uniform policy that submitted papers cannot be uploaded to repositories like arXiv (or made public in any way) starting one month from the submission deadline and extending through the time when decisions go out.
2. For specific conferences, check their sites for the precise date when this embargo goes into effect.
3. The policy is an attempt to balance the benefits of free and fast distribution of new ideas against the benefits of double-blind peer review.
4. For more on the policy and its rationale, see this ACL policy page.

# Typical NLP conference set-up

1. You submit your paper, along with area keywords that help determine which committee gets your paper.
2. Increasingly you also need to fill out very long and complicated checklists for various things the community cares about – try to find an expert to help you with this!
3. Reviewers scan a *long* list of titles and abstracts and then bid on which ones they want to do. The title is probably the primary factor in bidding decisions.
4. The program chairs assign reviewers their papers, presumably based in large part on their bids.
5. Reviewers read the papers, write comments, supply ratings.
6. Authors are allowed to respond briefly to the reviews.
7. The program/area chair might stimulate discussion among the reviewers about conflicts, the author response, etc.
8. The program committee does some magic to arrive at the final program based on all of this input. You might get a metareview that provides some insight into the final decision-making.

# Typical ACL set-up: Structured text

1. What is this paper about, what contributions does it make, and what are the main strengths and weaknesses?
2. Reasons to accept
3. Reasons to reject
4. Questions and additional feedback for the authors
5. Missing References
6. Typos, Grammar, Style, and Presentation Improvements
7. Ratings:
    a. Overall Recommendation
    b. Reviewer confidence
8. Confidential information (hidden from the authors)

# Author responses

Many conferences allow authors to submit short responses to the reviews. This is a rather uncertain business, but here are some thoughts:

1. Many people are cynical about author responses, since reviewers rarely actually change their scores afterwords.
2. It's bad in terms of signaling not to submit a response at all.
3. For conferences that have Area Chairs who are tasked with stimulating discussion and writing metareviewers for a small number papers, the author response might have a major impact.
4. NLP conferences often have complex rules about what you can and can't say in an author response. If you have questions about what you can do in a particular case, seek out an expert at Stanford for advice.
5. Always be polite. Be firm and direct, but do that strategically, to signal what you feel most strongly about.
    - **Never**: "Your inattentiveness is embarrassing; section 6 does what you say we didn't do."
    - **Yes**: "Thank you. The information you're requesting is in section 6. We will make this more prominent in our revision."

# Presentation types and venues

## Presentation types

- Oral presentations vs. poster presentations
- Workshops vs. main conferences

## Relevant conferences

- ACL
- NAACL
- EMNLP
- AACL
- EACL
- COLING
- CoNLL
- (Workshops)
- WWW
- WSDM
- KDD
- ICWSM
- IJCAI
- AAAI
- CogSci
- SCiL
- ICML
- NeurIPS
- ICLR

# My personal assessment of NLP reviewing

1. The focus on conference papers has been good for NLP. It fits with, and encourages, a rapid pace.
2. Before about 2010, the reviewing was admirably good and rigorous in comparison with other fields.
3. Lately, the growth of the field has reduced the general quality of reviewing; the field is still grappling with this.
4. Reviewers are occasionally *incredibly mean*. One needs to desensitize oneself to this. It can help to share your reviews with an experienced NLPer.
5. Forcing every paper to be 4 or 8 pages is not good, but this issue is being addressed productively with more use of supplementary materials.
6. The biggest failing: no chance for authors to appeal to an editor and interact with that editor. Journals allow this, to good effect.
7. *Transactions of the ACL* (TACL) is a journal that follows the standard ACL conference model fairly closely but allows for journal-style interaction with an editor.

# On titles

1. Jokey is risky*
2. Calibrate to the scope of your contribution
3. Consider the reviewers you are likely to attract
4. Avoid special fonts and formatting if possible

* Sagi, Yagi and Eldad Yechiam. 2008. Amusing titles in scientific journals and article citation. *Journal of Information Science* 34(5): 680–687.

# On abstracts

Important for creating a first impression. A general structure:

1. The opening is a broad overview – a glimpse at the central problem.
2. The middle takes concepts mentioned in the opening and elaborates upon them, probably by connecting with specific experiments and results from the paper.
3. The close establishes links between your proposal and broader theoretical concerns, so that the reviewer has an an answer to the question “Does the abstract offer a substantive and original proposal”.

# On abstracts

**Abstract**

*This opening sentence situates you, dear reader. Our approach seeks to address the following central issue: . . . The techniques we use are as follows: . . . Our experiments are these: . . . Overall, we find that our approach has the following properties: . . . (The significance of this is . . . )*

# On style sheets (or, on avoiding desk rejects)

1. Pay close attention to the details of the style-sheet and any other requirements included in the call for papers.
2. In NLP, infractions here are the most likely cause of the dreaded “desk reject” – rejection without review.

# The camera-ready version

1. “Camera-ready” refers to old-fashioned technology for publishing on paper!
2. For most NLP conferences, you get an additional page upon acceptance, presumably to respond to requests made by reviewers, though in practice you can use the space however you like.
3. In general, the extra page is probably used for fixing passages that were made overly terse in order to get the original submission within the required length limit.
4. You could also use it to improve your existing results, but very often substantially new ideas and results are better turned into a separate follow-up paper.

# Giving talks

# Basic structure

Mirrors paper structure, but **must be simpler**.

## Beginning

- What problem are you solving?
- Why is it important?
- What approaches have been tried, and why have they not fully solved the problem?

## Middle

- What data?
- What approach? (model type, feature representations)
- How to evaluate success?

## End

- Quantitative results, graphs.
- Which features/techniques/resources contributed most?
- What kinds of things do we still get wrong? Examples.
- Overall, what happened and why?

# Pullum’s Golden Rules

Geoff Pullum’s Five Golden Rules (well, actually six) for giving academic presentations:

1. Don’t ever begin with an apology.
2. Don’t ever underestimate the audience’s intelligence.
3. Respect the time limits.
4. Don’t survey the whole damn field.
5. Remember that you’re an advocate, not the defendant.
6. Expect questions that will floor you.

http://www.lel.ed.ac.uk/~gpullum/goldenrules.html

# Patrick Blackburn’s fundamental insight

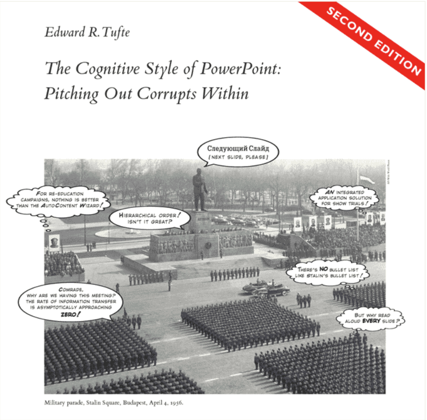
Where do good talks come from?

Honesty.

“A good talk should never stray far from simple, honest communication.”

https://web.stanford.edu/class/cs224u/readings/blackburn2001.pdf

# PowerPoint used for evil (not inevitable!)



SECOND EDITION

Edward R. Tufte

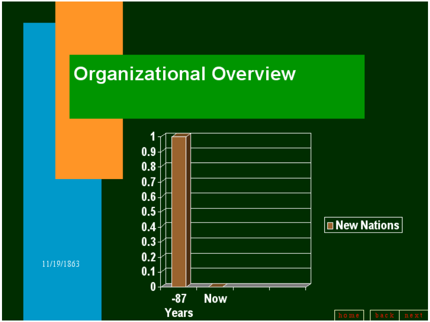
*The Cognitive Style of PowerPoint: Pitching Out Corrupts Within*

Следующий Слайд (NEXT SLIDE, PLEASE)
FOR RE-EDUCATION CAMPAIGNS, NOTHING IS BETTER THAN THE AUTOCONTENT WIZARD!
HIERARCHICAL ORDER! ISN'T IT GREAT?
AN INTEGRATED APPLICATION SOLUTION FOR SHOW TRIALS!
THERE'S NO BULLET LIST LIKE STALIN'S BULLET LIST!
COMRADE, WHY ARE WE HAVING THIS MEETING? THE RATE OF INFORMATION TRANSFER IS ASYMPTOTICALLY APPROACHING ZERO!
BUT WHY READ ALOUD EVERY SLIDE?

Military parade, Stalin Square, Budapest, April 4, 1956.

http://www.edwardtufte.com/tufte/powerpoint

Peter Norvig: Gettysburg Address as PowerPoint



Organizational Overview

http://norvig.com/Gettysburg/

# Slide design: two schools of thought

## Minimalist

1. Your slides should be as spare as possible.
2. The audience should spend most of the time listening to and looking at you.
3. Individual slides do not stay up for long or get used in more than one way.

## Comparative

1. Your slides should be as full as possible without sacrificing clarity.
2. Your talk should make it easy for people to spend time studying your slides.
3. Individual slides stay up for a long time and get used to make multiple comparisons and establish numerous connections.

# Slide design: two schools of thought

## A personal matter

# Slide design: two schools of thought

## A personal matter

- The minimalist view seems right for telling a story – often the best mode when time is of the essence and the audience is mainly there to learn about what your paper contains.

# Slide design: two schools of thought

## A personal matter

- The minimalist view seems right for telling a story – often the best mode when time is of the essence and the audience is mainly there to learn about what your paper contains.
- The comparative view seems right for teaching; it’s the closest slides come to a full, well-organized chalkboard.

# Slide design: two schools of thought

## A personal matter

- The minimalist view seems right for telling a story – often the best mode when time is of the essence and the audience is mainly there to learn about what your paper contains.
- The comparative view seems right for teaching; it's the closest slides come to a full, well-organized chalkboard.
- Find the style that works for you. As long as you think long and hard about what it will be like to listen to your talk, and adjust accordingly, you'll shine.

# Guiding audience attention

# Guiding audience attention

1. Use overlays to fill a slide while still keeping the audience with you.

# Guiding audience attention

1. Use overlays to fill a slide while still keeping the audience with you.

2. Color used **systematically** to create distinctions.

# Guiding audience attention

1. Use overlays to fill a slide while still keeping the audience with you.

2. Color used **systematically** to create distinctions.

3. Size to draw attention to things.

# Guiding audience attention

1. Use overlays to fill a slide while still keeping the audience with you.
2. Color used **systematically** to create distinctions.
3. Size to draw attention to things.
4. Boxes, arrows ←, and other devices to help people navigate plots, model diagrams, and long prose statements.

# Guiding audience attention

Overlays

# Guiding audience attention

Color

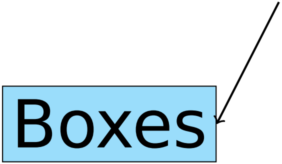
# Guiding audience attention

Size

# Guiding audience attention

Boxes

# Guiding audience attention



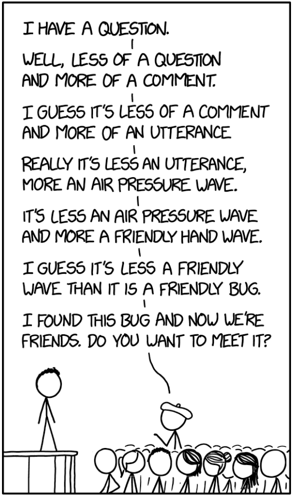
Boxes

# More mundane things

- Turn off any notifications that might appear on the screen.
- Make sure your computer is out of power-saver mode so that the screen doesn’t shut off while you’re talking.
- Shut down running applications that might get in your way.
- Make sure your desktop is clear of files and notes that you wouldn’t want the world to see.
- If using PowerPoint / Keynote / Google Slides, have a PDF back-up just in case.
- Projectors can fail; always be prepared to give the talk without slides.

# The discussion period

1. This is an important part of the presentation.
2. It should be a chance for the audience to gain a deeper understanding of your ideas. When the entire discussion period has this aim, it is a joy.
3. But sometimes other things happen: hostile questioners, confused questioners, . . .
4. Try to pause for one second before answering each question.
5. Avoid saying “I have no idea” and leave it at that. When floored, say: “I have no idea, but let’s think about . . . ”
6. Most questions won’t make total sense to you. Your questioner doesn’t know the work all that well.
7. You’ll be a hit if you can warp every question you get into one that makes sense and leaves everyone with the impression that the questioner raised an important issue.


I HAVE A QUESTION.
WELL, LESS OF A QUESTION AND MORE OF A COMMENT.
I GUESS IT'S LESS OF A COMMENT AND MORE OF AN UTTERANCE
REALLY IT'S LESS AN UTTERANCE, MORE AN AIR PRESSURE WAVE.
IT'S LESS AN AIR PRESSURE WAVE AND MORE A FRIENDLY HAND WAVE.
I GUESS IT'S LESS A FRIENDLY WAVE THAN IT IS A FRIENDLY BUG.
I FOUND THIS BUG AND NOW WE'RE FRIENDS. DO YOU WANT TO MEET IT?

https://xkcd.com/2191/

# References I

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumeé III, and Kate Crawford. 2018. Datasheets for datasets. *arXiv preprint arXiv:1803.09010*.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pages 220–229, New York, NY, USA. Association for Computing Machinery.

Priyanka Nanayakkara, Jessica Hullman, and Nicholas Diakopoulos. 2021. Unpacking the expressed consequences of ai research in broader impact statements. *arXiv preprint arXiv:2105.04760*.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237. Association for Computational Linguistics.