Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Reviewing in NLP – A Tutorial

Kevin Bretonnel Cohen, Karën Fort, Margot Mieskes,
Aurélie Névéol

July 5th, 2020

1

How to Review specific Sections in a Scientific Paper

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

1 Introduction

2 Background

3 The Actual Review

4 Ethics and reviewing

5

Who are we?
Kevin

Who are we?
Karën Fort

Reviewing in NLP – A Tutorial

Kevin Bretonnel Cohen, Karën Fort, Margot Mieskes, Aurélie Névéol

Introduction

Background

The Actual Review

Ethics and reviewing

- Expertise in language resources for NLP
- Reviewing Approach: "Write it the way you'd like to read it for yourself"
- Biases:
    - "[...] everything looks like a nail" (i.e. I'm a specialist in manual annotation, so I tend to focus too much on it)
    - Strong preference for well-written papers
    - . . .

4

Who are we?
Margot Mieskes

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- Summarization, Evaluation, Replication, Ethical Concerns in NLP
- Beating state-of-the-art is not the main goal
- Biases:
  - Higher, Faster, Better is not Science
  - What can be learned about the use of language
  - ...

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Who are we?
Aurélie Névéol

- Expertise in biomedical NLP
- Reviewing Approach: "Come with an open mind"
- Biases:
  - Strong preference for IMRaD structured papers
  - Interest in evaluation beyond performance measures
  - ...

6

Margot Starts Here!

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Introduction

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Welcome

Starting Point

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Dear author, we are sorry to inform you that your paper has not been accepted for publication at <INSERT PUBLICATION VENUE>.

Starting Point

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Dear author, we are sorry to inform you that your paper has not been accepted for publication at <INSERT PUBLICATION VENUE>.

https://upload.wikimedia.org/wikipedia/commons/thumb/4/42/Sad_smiley_yellow_simple.svg/
1024px-Sad_smiley_yellow_simple_svg.png

9

And then the details ...

Reviewing in NLP – A Tutorial

Kevin Bretonnel Cohen, Karën Fort, Margot Mieskes, Aurélie Névéol

Introduction

Background

The Actual Review

Ethics and reviewing

This is a bad paper and should not be published.

These examples are quoted from real reviews (or slightly reworded); some quotes are also discussed on
https://hackingsemantics.xyz/2020/reviewing-models/

And then the details ...

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

"The main weakness of the paper is the results do not beat the
state of the art models."

And then the details ...

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

There is nothing really new about this paper.
"This paper presents yet another question answering test."

And then the details ...

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

"The paper is mostly a description of the corpus and its collection and contains little scientific contribution."

And then the details ...

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

"I personally do not care for this topic."

And then the details ...

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

"This must be the work of an inexperienced student".

And then the details ...

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

"Not so good English."

And then the details ...

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

https://upload.wikimedia.org/wikipedia/commons/thumb/4/42/Smiley_green_alien_unhappy.svg/
807px-Smiley_green_alien_unhappy_svg.png

Problems with reviewing

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- ACL 2019 Survey Results
- A lot of discussion on
  - how to deal with review load
  - how to lower the review load
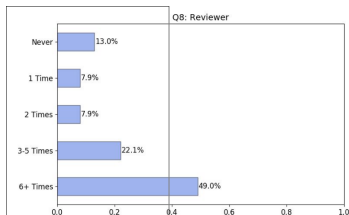- Criticism on Quality of Reviews
- Demographic information about reviewers

Source:
http://acl2019pcblog.fileli.unipi.it/wp-content/uploads/2019/07/ReportACL2019ReviewingSurvey.pdf

Problems with reviewing

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Results included:

- General Comments: Concerns about reviewer preparation, review quality, and reviewing load; ...
- Structure Forms improve Quality/Clarity
- Having at least some positive and some negative points provides a more balanced feel and is important
- Can help reduce bias and increase fairness in judgements
- Structured forms are helpful for new reviewers

Source:
http://acl2019pcblog.fileli.unipi.it/wp-content/uploads/2019/07/ReportACL2019ReviewingSurvey.pdf

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Problems with reviewing

Results included:

- General Comments: Concerns about reviewer preparation, review quality, and reviewing load; ...
- Structure Forms improve Quality/Clarity
- Having at least some positive and some negative points provides a more balanced feel and is important
- Can help reduce bias and increase fairness in judgements
- Structured forms are helpful for new reviewers

Source:
http://acl2019pcblog.fileli.unipi.it/wp-content/uploads/2019/07/ReportACL2019ReviewingSurvey.pdf

Problems with reviewing

Reviewing in NLP – A Tutorial

Kevin Bretonnel Cohen, Karën Fort, Margot Mieskes, Aurélie Névéol

Introduction

Background

The Actual Review

Ethics and reviewing

Results included:

Reviewer quality is a major concern / training for new reviewers might be useful

Source:
http://acl2019pcblog.fileli.unipi.it/wp-content/uploads/2019/07/ReportACL2019ReviewingSurvey.pdf

Therefore

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

That's why we are here!

Switch/Break

Reviewing in NLP – A Tutorial

Kevin Bretonnel Cohen, Karën Fort, Margot Mieskes, Aurélie Névéol

Introduction

Background

The Actual Review

Ethics and reviewing

source: https://de.wikipedia.org/wiki/Datei:Mt._Lavinia-_Governor%27s_High_Tea.jpg

Exercise: Which of the following makes a stronger significance argument?

Reviewing in NLP – A Tutorial

Kevin Bretonnel Cohen, Karën Fort, Margot Mieskes, Aurélie Névéol

Introduction

Background

The Actual Review

Ethics and reviewing

https://commons.wikimedia.org/wiki/File:Man_Doing_Warm_Up_Exercise_Cartoon.svg

1. There has recently been considerable interest in the problem of compositionality of embeddings.
2. There has recently been considerable interest in the problem of compositionality (Able 2015).
3. There has recently been considerable interest in the problem of compositionality of embeddings (Able 2015, Baker 2016, Charlie 2017, Delta 2018, Echo 2019).

14

Exercise: Which of the following makes a more actionable reviewer comment?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

https://commons.wikimedia.org/wiki/File:Man_Doing_Warm_Up_Exercise_Cartoon.svg

1. The author must be a student who was not taught how to properly support a claim.

2. The statement "There has recently been considerable interest in the problem of compositionality" is unsupported.

3. The following references could be used to support the claim "There has recently been considerable interest in the problem of compositionality": (Able 2015, Baker 2016, Charlie 2017, Delta 2018, Echo 2019).

15

Background

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Exercise: Which of the following provides the best support for a conclusion that makes a claim of state-of-the-art performance against a baseline of 0.95?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

1. Average F-measure as obtained from cross-validation was 0.96.

2. Median F-measure as obtained from cross-validation was 0.96.

3. Median F-measure as obtained from cross-validation was 0.96 with an inter-quartile range of 0.4, minimum of 0.92, and maximum of 0.96.

4. Median F-measure as obtained from cross-validation was 0.94 with an inter-quartile range of 0.02, minimum of 0.91, and maximum of 0.96.

https://commons.wikimedia.org/wiki/File:Man_Doing_Warm_Up_Exercise_Cartoon.svg

Why do reviews?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Why do reviews?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- For the good of the World
- For the good of the research community
- For your own good

Who is your audience?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Who is your audience?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- The authors of the paper:
  - *narrative feedback* helps authors understand what the paper conveys
  - *actionable feedback* helps improve the paper
- The Area Chairs / Editor: *itemize major pros and cons* to help reach a decision
  - *performance point of view* helps them evaluate the novelty/correctness of the research
  - *deontological point of view* helps them evaluate the way the research was conducted

Exercise

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

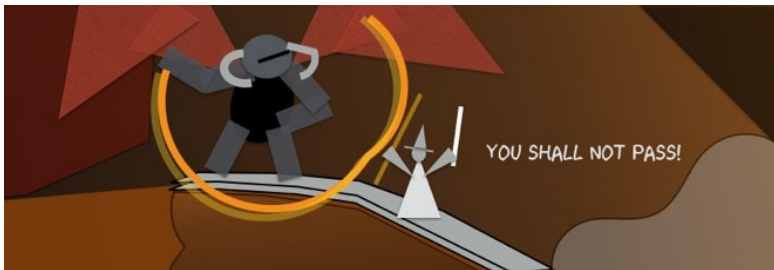The Actual
Review

Ethics and
reviewing

https://commons.wikimedia.org/wiki/File:Man_Doing_Warm_Up_Exercise_Cartoon.svg

You are reviewing for a high-status conference, and for a low-status conference. How should your reviews differ?

Exercise

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction
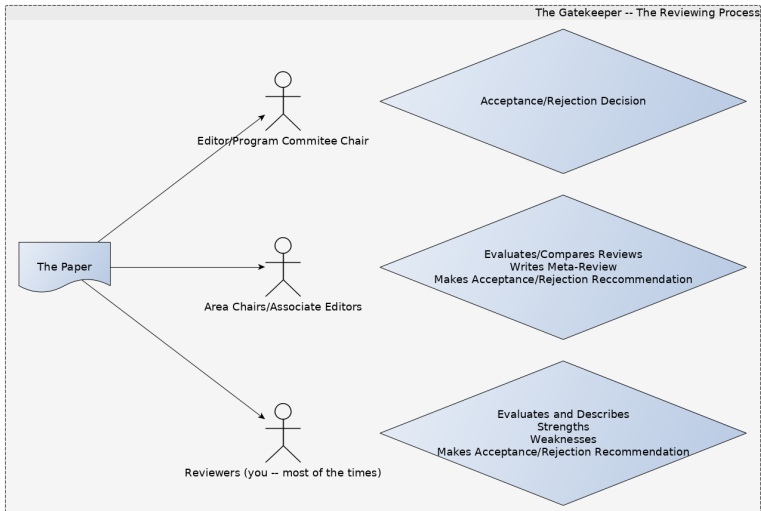
Background

The Actual
Review

Ethics and
reviewing

https://commons.wikimedia.org/wiki/File:Man_Doing_Warm_Up_Exercise_Cartoon.svg

You are reviewing for a high-status conference, and for a low-status conference. How should your reviews differ?

Why?

Roles

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
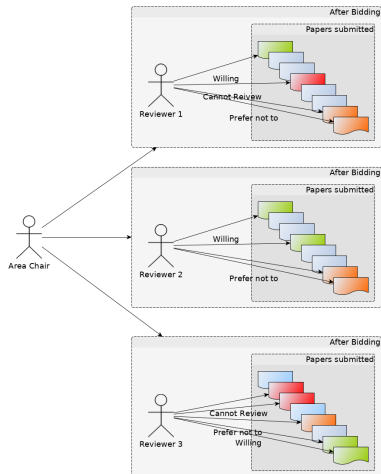Review

Ethics and
reviewing

The Gatekeeper

https://www.wired.com/2014/08/lotr-physics/

Roles

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

The Gatekeeper

https://www.wired.com/2014/08/lotr-physics/
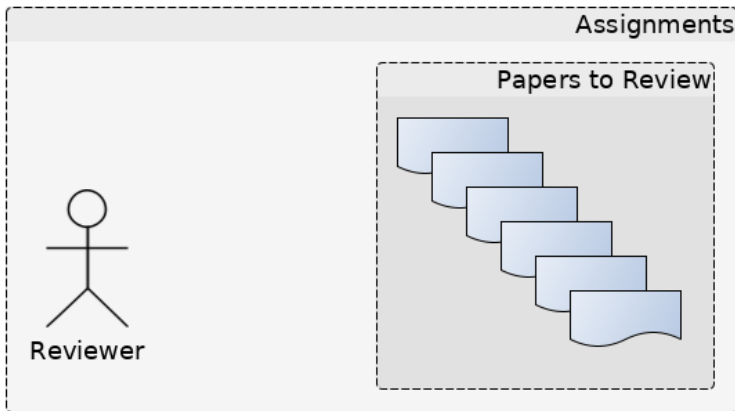
Not (only) YOU!

Roles

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

The gatekeeper:

How do they get to you?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Bidding

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing
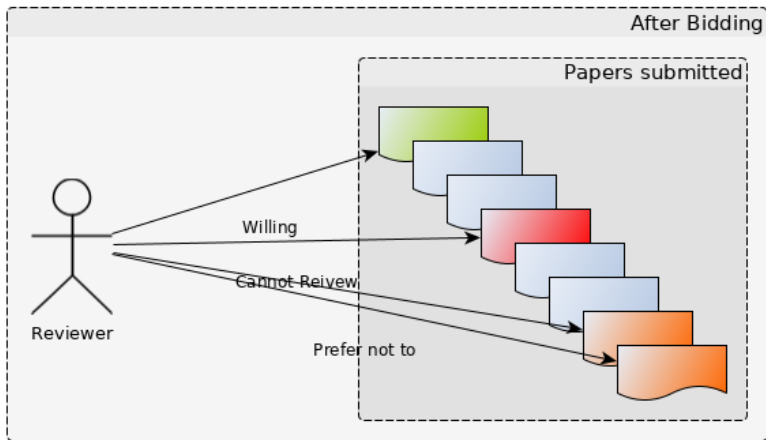
How do they get to you?

Assignment

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
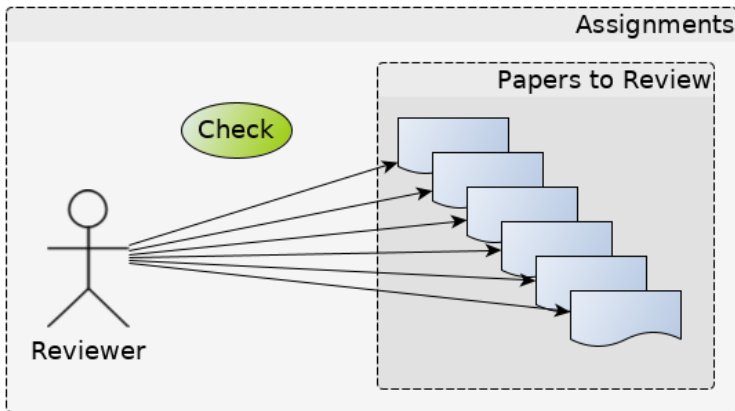Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Receiving Assignments

How do they get to you?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Check Papers

How do they get to you?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- Conflict of Interest?
- Concerns about your ability?
    - Topic is somehow offensive?
    - Methodology is outside of your skill scope
    - . . .

Inform the Area Chair ASAP.

What happens next?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Author Response Period.

https://www.pexels.com/photo/questions-answers-signage-208494/

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Discussion Phase

https://freesvg.org/discussion

Exercise: What problem – if any – do you see with this fragment of a review?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

https://commons.wikimedia.org/wiki/File:Man_Doing_Warm_Up_Exercise_Cartoon.svg

Quantitative score: *Methodology: 1 (Methods are invalid for testing the hypothesis)*

Exercise: What problem – if any – do you see with this fragment of a review?

Reviewing in
NLP – A
Tutorial
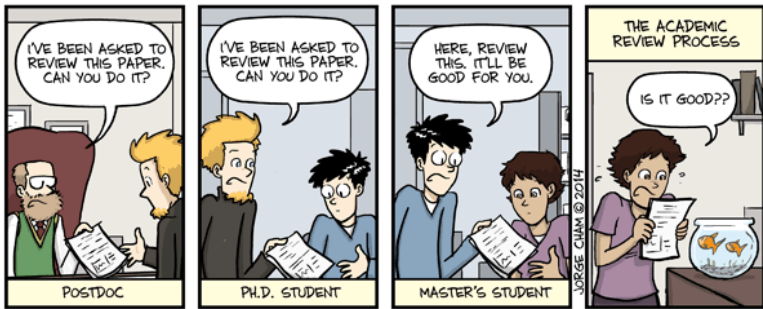
Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

https://commons.wikimedia.org/wiki/File:Man_Doing_Warm_Up_Exercise_Cartoon.svg

Comments to authors: *The system uses rules. But, machine learning is cheaper than rules.*

25

Margot Ends Here

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Switch/Break

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

source: https://de.wikipedia.org/wiki/Datei:Mt._Lavinia-_Governor%27s_High_Tea.jpg

Karen Starts Here

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

The actual review

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Source: http://phdcomics.com/comics.php?f=1760

29

Exercise: Reviewing the reviews

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- Read **Paper 42** (optional)
- Read the reviews provided with **Paper 42**
- Do you think they are good/bad reviews? Why?
- Live discussion on July 5th

What makes a ~~good~~ bad (easier to tell) review?
From the authors' point of view

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- superficial reading of the paper

- rudeness or nastiness

- inflexibility

- expecting perfection

- reviewing your version of the paper rather than the submitted paper: interest in the topic, feeling about the paper, . . .

- making unsupported claims about the paper

- [insert yours here]

Sources: https://acl2017.wordpress.com/2017/02/23/last-minute-reviewing-advice/ and
https://sites.umiacs.umd.edu/elm/2016/02/01/mistakes-reviewers-make/

What makes a ~~good~~ bad (easier to tell) review?
From the area chairs/editors' point of view

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- a late review (without advance warning and planning)
- a lukewarm review
- a positive review with low scores (an incoherent review)
- an insulting review
- [insert yours here]

Sources: https://acl2017.wordpress.com/2017/02/23/last-minute-reviewing-advice/ and
https://coling2018.org/reviewer-code-of-conduct/

Exercise: Writing a review

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- Read **Paper 42**
- Write a <u>short</u> review
- Live discussion on July 5th

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

What you should ask yourself

- Did I take enough time to digest the paper?
- Did I take enough time to write/review my review?
- Did I focus on the work?
- Are the comments I made constructive?
- Are the comments I made substantial enough (are they grounded)?
- Did I mention something good about the paper?
- Did I check:
  - the accompanying code/data (if any)?
  - potential (self-)plagiarism? Salami-slicing?

?? How would I feel about receiving such a review?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

How did you proceed?
One possibility. . .
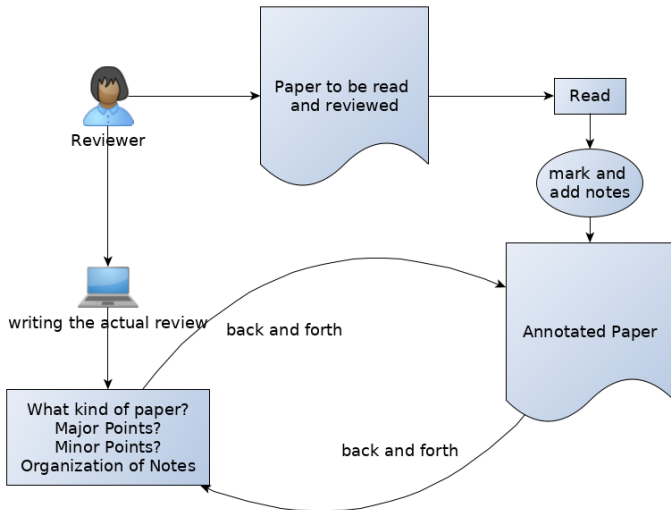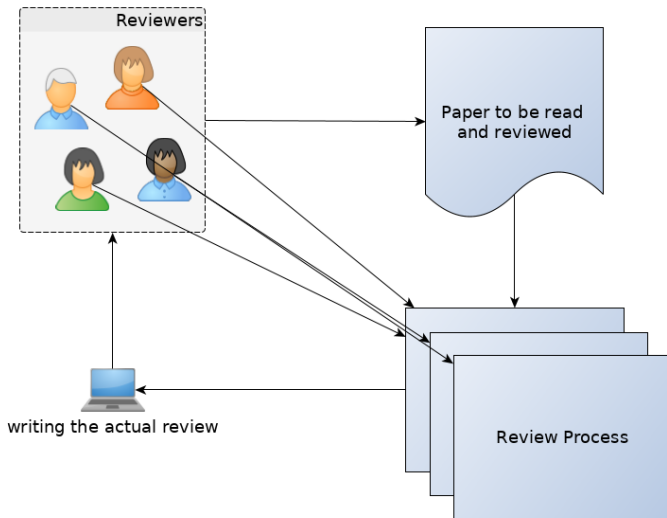
🔍 Reading is annotating

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

🔍 Approach

- What type of paper is it?
  - Method paper
  - Evaluation Paper
  - Reproducibility paper
  - Resource Paper
  - Position/survey paper

- Identify the paper type and use appropriate evaluation criteria

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

NLP engineering experiment paper

Example from http://coling2018.org/paper-types/

- Can you tell what research question was addressed?
  - What is the hypothesis?
  - How was it tested?
- Is the research technically sound and correct?
- Is there an error analysis?
- Is there meaningful comparison?
  - To the literature?
  - To baselines?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Reproduction paper
Example from http://coling2018.org/paper-types/

- Data/Code availability
    - Of the original research
    - Of the reproduction
- Description of the specific reproduction obstacles
- Generalizability: beyond reproduction?
- Informativeness: insight of the reproduction?
- Is there meaningful comparison to the original experiments?

🔍 Resource paper

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- How was the resource produced?
    - Source material?
    - Workforce?
- How was the resource evaluated?
    - What was the quality control procedure?
    - Was inter-annotator agreement computed?
- Are metadata and accessibility detailed?
- How does it compare to existing resources?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

🔍 Survey paper

Karen: Final slide for everybody's usage (to improve)

- Organization: does the survey address a specific question?
- Scope: Does the paper identify a reasonably focused area to survey?
- Thoroughness: does the paper cover all of the relevant literature?
  - Is the source of the literature surveyed clear?
  - Are criteria for inclusion/exclusion of papers in the survey clear?
- Outlook: Does the paper identify areas for future work ?

42

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

🔍 Notes

- What is the paper about?
- What do they use?
  - Data
  - Methods
  - Tools
  - . . .
- What are the results?
- What are the claims?
- Do results and claims match?

First (or last?) step: write a summary!

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

🔍 Notes

- Are the hypotheses sound?
- Is the experimental setup suited to the problem stated?
- Is the evaluation sound?
- Is the problem well-motivated?
- What is the take-home message?
- Are the conclusions supported by the results?
- . . .

🔍 Writing the review

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- Follow the order of the paper
- List strengths and weaknesses (?)
- Major *vs.* minor issues

  Support your claims with appropriate references

Switch/Break

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

source: https://de.wikipedia.org/wiki/Datei:Mt._Lavinia-_Governor%27s_High_Tea.jpg

45

Most conferences in the domain use the same basic type of
form with:

- numeric scores
- free text form(s)
- confidential comments

Exercise: What problem do you see with the following fragment of a review?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- **Quantitative score:** Review of related literature: 1 (Major relevant papers are missing)
- **Comments to authors:** There are missing references.

Exercise: What problem do you see with the following fragment of a review?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- **Quantitative score:** Review of related literature: 1 (Major relevant papers are missing)
- **Comments to authors:** There are missing references.

How could it be improved?

47

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

🔍 Numeric scores (beware!)

- Consistent with your textual review
- Sanity Check

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Online *vs* offline review forms

or else. . .

<div>

Online

+ easy to fill in

- limited session (you might
  lose your scores/text)

</div>

<div>

Offline

+ no limit in time

- not so easy to fill in
  (old-school text format)

</div>

Discussion clipart: MRafizeldi / CC BY-SA

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Online *vs* offline review forms

or else. . .

Online	Offline
+ easy to fill in	+ no limit in time
- limited session (you might lose your scores/text)	- not so easy to fill in (old-school text format)
- bias due to the form?	- bias due to the form?

$\rightarrow$ (Possible) Mixed approach: write your (text) review in a
text file, then copy/paste in the online form

Discussion clipart: MRafizeldi / CC BY-SA

49

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

ACL 2020 review form

- "What is this paper about, what contributions does it make, what are the main strengths and weaknesses?"
- "Reasons to accept"
- "Reasons to Reject"
- "Overall Recommendation"
  - "Transformative"
  - "Exciting"
  - "Strong"
  - "Leaning Positive"
  - "Leaning Negative"
  - "Mediocre"
  - "Weak"
  - "Poor"

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

ACL 2020 review form

- "Reviewer Confidence"
- "Author Response"
- "Questions for the Author(s)"
- "Missing References"
- "Typos, Grammar, Style"
- "Additional Suggestions"

- "Confidential Information"

"What is this paper about, what contributions does it make, what are the main **strengths** *and* **weaknesses***?"*
[*ACL 2020 review form*]

vs

Major *points and* **minor** *points.*
[*Most bio-info review forms*]

Discussion clipart: MRafizeldi / CC BY-SA

Using confidential Comments

Source: Free photo by Tophee Marquez from https://www.pexels.com

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

🔍 Using confidential Comments

- Additional Comments wrt. the overall recommendation
- Special circumstances that apply to this paper and might allow for acceptance/rejection despite low(er)/high(er) marks
- Comments about the author(s) rather than the paper
- Plagiarism / salami slicing concerns
- Concerns about your own review (level of expertise, discovery of COI...)
- Anonymity breaches (including for *you*: perhaps you have reviewed the same paper before...)

Anything else **you** think is important, but should not be visible to the other reviewers and/or the authors

(very) Bad papers?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Do not waste your time

vs

make a pedagogical effort

Discussion clipart: MRafizeldi / CC BY-SA

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

(very) Bad papers?

Do not waste your time

vs

make a pedagogical effort → depends if you feel like the

authors made an effort?

Discussion clipart: MRafizeldi / CC BY-SA

What else?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Plagiarism

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

source: https://www.merriam-webster.com/dictionary/plagiarizing
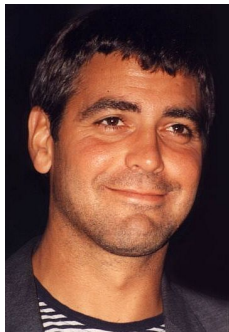
Exercise: Back to Paper 42

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- Re-read _Paper_42
- Is it possible that part of it is, in fact, plagiarism?
- If yes, find it and find the source

From plagiarism. . .

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Plagiarism detection software is usually available in your
University, e.g. Compilatio (Sorbonne University), iThenticate
(University of Amsterdam), etc. . .

Or you have your own similarity detection system

Or your favorite search engine

Just use it!

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

. . . to salami-slicing detection

A side effect of looking for plagiarism is that you may find salami-slicing (or self-plagiarism):

> "Salami publication (sometimes called bologna or trivial publication) is the practice of dividing one significant piece of research into a number of small experiments [...] to increase the number of publications."

https://www.law.hku.hk/researchintegrity/salami-slicing/

Exercise: Research material in Paper 42

- Is the material available?
- Does it look relevant?

Reproducibility

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

When data is provided or linked to:

- check if it really exists
- check if it does not break the anonymity
- if time allows, try to download and run it

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

🔍 Writing a review

- Plan enough time to do it
- Be polite
- Focus on the work
- Avoid trivialities
- Do not attack the authors
- Take specifics into account (e.g. low-resourced language)
- Be as elaborate and explicit as possible
- The worse the numeric scores, the more elaborate the review should be to support this

Switch/Break

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

source: https://de.wikipedia.org/wiki/Datei:Mt._Lavinia-_Governor%27s_High_Tea.jpg

Karen Ends Here

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Aurelie Start Here

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Ethics in reviewing

A two way street:

1. Provide fair, balanced, ethical reviews
2. Hold authors accountable for their research

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Introductory story

source: https://www.bbc.com/news/amp/science-environment-34910954

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Introductory story

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Introductory story

Why this statement?

What's the first explanation to pop into your head?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- Some answers:
  - The son is adopted (biological father and adoptive father)
  - Same sex parents (two fathers)
  - The surgeon is also a catholic priest
  - ...
  - Wait! Could the surgeon be... the MOTHER?

- Why isn't the most simple answer obvious to everyone?
  - **Implicit Bias: surgeon = man**

67

Implicit bias

"*Bias that results from the tendency to process information based on unconscious associations and feelings, even when these are contrary to one's conscious or declared beliefs*"

It doesn't only happen to other people!

Reviewing in
NLP – A
Tutorial
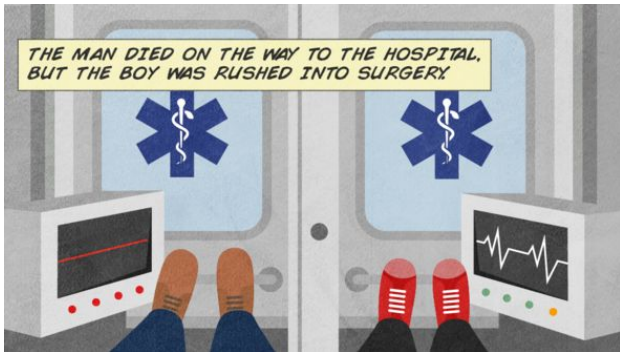
Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Project Implicit®

- Project Implicit
  - A research project that investigates social processes involved in social interactions
  - Covers many aspects including social attitudes, health
  - Offers a set of "Implicit Association Tests" (IAT) - free for non commercial use

Exercice: take a test

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

https://implicit.harvard.edu/implicit/takeatest.html

- Suggested: Gender-Science (will be randomly positioned in left side menu)

- Requires about 10/15 minutes, including filling preliminary question forms and test

- Measures reflex associations between words, ideas, pictures

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Results of a sample test: Gender-Science IAT

Percent of web respondents with each score

Strong automatic association of Male with Science and Female with Liberal Arts — 23%

Moderate automatic association of Male with Science and Female with Liberal Arts — 29%

Slight automatic association of Male with Science and Female with Liberal Arts — 18%

Little to no automatic preference between gender and academic domains — 19%

Slight automatic association of Male with Liberal Arts and Female with Science — 6%

Moderate automatic association of Male with Liberal Arts and Female with Science — 4%

Strong automatic association of Male with Liberal Arts and Female with Science — 1%

This distribution summarizes 628,295 IAT scores for the Gender-Science task completed between January 2003 and December 2015.

source: *Xu, Kaiyuan et al. "Project Implicit Demo Website Datasets." OSF, 17 Mar. 2020. Web.*
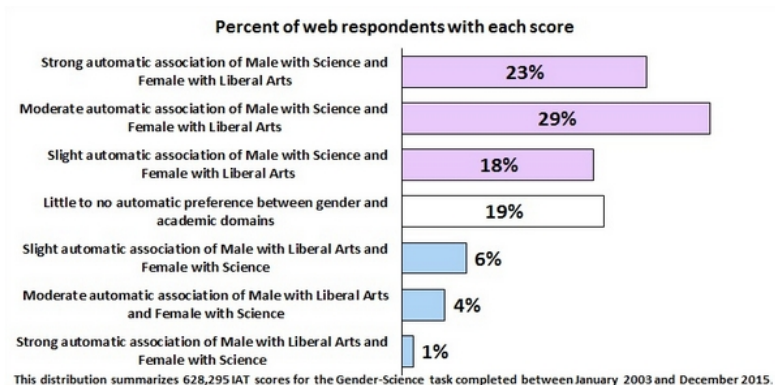
Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

How can we handle our biases?
An open question...

⚖ + Bias awareness can induce guilt and motivation to change behaviour

Parker LR, Monteith MJ, Moss-Racusin CA, Van Camp AR, Promoting concern about gender bias

with evidence-based confrontation, J Exp Soc Psychol. 2018;74:8-23

⚖ - Stereotype prevalence can create bias reinforcement (everybody else is doing it, so why can't we?)

Duguid MM, Thomas-Hunt MC. Condoning stereotyping? How awareness of stereotyping

prevalence impacts expression of stereotypes. J Appl Psychol. 2015;100(2):343-359.

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

How can we handle our biases?
An open question...

Qualify your comments where applicable "I am [not] an expert on topic XYZ, and here is my take"

🔍 Biases that may affect reviewing

- Gender
- Seniority
- Geography/institution
- Reputation

```
Kaatz A, Gutierrez B, Carnes M. Threats to objectivity
in peer review:  the case of gender.  Trends in
pharmacological sciences.  2014;35(8):371-373.

Okike K, Hug KT, Kocher MS, Leopold SS. Single-blind
vs Double-blind Peer Review in the Setting of Author
Prestige.  JAMA. 2016 Sep 27;316(12):1315-6.
```

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Reviewing models: a balancing act
reviewer accountability *vs.* bias effect

Open Review Papers, Reviews and Identities are visible to all
parties

Single-Blind Author is known to Reviewer – Reviewer remains
unknown to Author

Double-Blind Author is unknown to Reviewer and Reviewer is
unknown to Author

Pros and Cons

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

open

- Reviewer is accountable
- Authors do not need to withhold scientific content to preserve anonymity
- Reviewer may withhold criticism or face retaliation
- Vulnerable to all biases

Pros and Cons

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

single-blind

- Allows the reviewer to be fully honest
- Authors do not need to withhold scientific content to preserve anonymity
- Vulnerable to all biases mentioned above

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Pros and Cons

double-blind

- Avoids various biases
- Allows the reviewer to be fully honest
- Reviewer is not accountable
- Authors may withhold scientific content to preserve anonymity

Holding the reviewers accountable

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

- AC/editor's role in the reviewing process
- Secondary reviewers should be declared
- Feedback mechanisms are essential

Holding the authors accountable

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- Check for (self-)plagiarism
- Check for salami slicing
- Check for reproducibility material
- Check for ethical concerns
    - data confidentiality
    - research process

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Organizing reviewing and ethics
Making the sausage

- How do you define and handle conflicts of interest??
  - Are organizers allowed to submit?
- Who should see reviewers' names?
- Organizing the discussion between reviewers
- Transparency of the process:
  - AC names are mentioned somewhere
  - Reviewing process is detailed somewhere
- Rebuttal period / Authors' response
- Meta reviews

78

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Characteristics of Various Venues
(with changes over the years)

Venue	Blind	Supplements / appendix	Area Chair	Response Period
*ACL	Double	yes	identified	sometimes
EMNLP	Double	yes	identified	yes
IJCAI	Double	no	internal	yes
AAAI	Double	no	internal	yes
LREC	Single	no	no	no
Swisstext	Double	no	internal	no
TALN	Double	no	internal	no
AMIA	Single	no	internal	no

Closing remarks
Reviewing the review tutorial

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

- It's been a bumpy ride
- Thank you for your attendance
- Your feedback is welcome
- See (hear) you during the live session!

We would love to hear from you!

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

kevin.cohen@gmail.com
karen.fort@sorbonne-universite.fr
margot.mieskes@h-da.de
aurelie.neveol@limsi.fr

81

Aurelie End Here

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Kevin Start Here

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

How to evaluate an innovation claim

Three most likely possibilities:

1. There is a well-supported innovation claim
2. There is a poorly-supported innovation claim
3. There is no innovation claim, or only an implied one

What "innovation" means in a research context

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

General structure of an innovation claim

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

If you cannot find anything like this structure, there might not actually be an explicit innovation claim in the paper.

Some ways for research to be innovative

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Extra slide: The difference between "innovation" and "significance"

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Exercise: Which of the following makes the strongest innovation claim?

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Introduction

Background

The Actual
Review

Ethics and
reviewing

Karen: I don't see the link with what is before, it should probably be in the innovation part (KBC)

1. *Able (2016) used a linear model. Baker (2017) used a linear mixed effects model. But, no one has used a support vector machine.*

2. *Able (2016) used a linear model. Baker (2017) used a linear mixed effects model. Charlie (2017) used a linear mixed effects model with post-processing. Delta (2017) used a linear mixed effects model with pre-processing. But, no one has used a support vector machine.*

3. *Able (2016) used a linear model. Baker (2017) used a mixed effects linear model. Charlie (2017) used a linear mixed effects model with post-processing. Delta (2017) used a linear mixed effects model with pre-processing. But, no one has*

89

Exercise: Which of the following makes the strongest innovation claim?

Reviewing in NLP – A Tutorial

Kevin Bretonnel Cohen, Karën Fort, Margot Mieskes, Aurélie Névéol

Introduction

Background

The Actual Review

Ethics and reviewing

Karen: I don't see the link with what is before, it should probably be in the innovation part (KBC)

1. *Able (2016) used a linear model. Baker (2017) used a linear mixed effects model. But, no one has used a support vector machine.*

2. *Able (2016) used a linear model. Baker (2017) used a linear mixed effects model. Charlie (2017) used a linear mixed effects model with post-processing. Delta (2017) used a linear mixed effects model with pre-processing. But, no one has used a support vector machine.*

3. *Able (2016) used a linear model. Baker (2017) used a mixed effects linear model. Charlie (2017) used a linear mixed effects model with post-processing. Delta (2017) used a linear mixed effects model with pre-processing. But, no one has*

Kevins Material

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol

Kevin End Here

Reviewing in
NLP – A
Tutorial

Kevin
Bretonnel
Cohen, Karën
Fort, Margot
Mieskes,
Aurélie Névéol