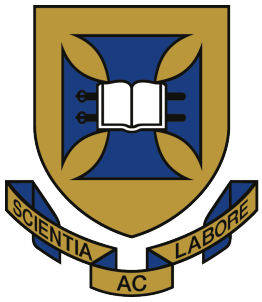

THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

ENGG7811: Research Methods

Assignment 1

Benjamin Greenland 46416403

Course Coordinator : Prof. Janet Wiles

September 2, 2024

Contents

List of Tables

1 Referencing Managers

a) For this assignment, Overleaf was used as a LaTeX text editor, and Zotero 7.0.2 was used as the referencing manager. I could not get latex to properly format the reference list (required for Task 6), resulting in the authors not appearing for some references (see Appendix B). To resolve this the reference list was made in Word and loaded into the LaTeX document.

b) I will be using cite-as-you-write when preparing this assignment as it reduces the risk of missing citations and makes it easier to manage references. It also ensures the sources are appropriate before building upon their information.

2 References & Bibliographic Styles

a) According to the arXiv.org website [1], arXiv.org is an open access repository that helps spread scholarly research by providing several services, including: "article submission, compilation, production, retrieval, search and discovery, web distribution for human readers, and API access for machines". Currently, the disciplines that deposit articles to arXiv.org include: "physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics" [1].

b) Submissions to arXiv.org are not peer reviewed, however they do pass through a moderation process to determine the submission's scholarly value and to classify it into a subject field. As a result, it is "the responsibility of the submitter" [1] to ensure the contents of each submission are appropriate.

In contrast, submissions to an online conference proceedings repository, like neurips.cc, are typically submitted to a review process [2]. This means the submissions have been reviewed by experts in the submission's topic. They are reviewed to ensure the use of proper research methods, and to assess the significance of the submission's contribution to existing literature [3].

This is based on information found on the arXiv.org website, as well as information found on the University of Notre Dame Australia and American Public University Library websites.

c) The main difference between the type of articles included in citation databases like Google Scholar and Web of Science is where the the original article was published, its format and the type of information it provides. This informs the review process the article may have gone through, as well as the depth of information the article provides. Typically, journal articles are the most credible in-depth sources as they contain the full research (including literature review, methodology, result, analysis, etc.) and are typically peer reviewed. Conference publications are similar to journal articles, however they are a more

condensed version to be presented at a conference. Preprints are also similar to journal articles but are not peer reviewed. In contrast, an editorial letter typically contains the opinions of a person or group on a published work [4].

d) This is the in-text citation for the preprint version [5].

This is the in-text citation for the online publication version [6].

e) The number of pages for each version of the paper is recorded below (Table 1).

Table 1: Number of pages in each version of the paper.

Article Version	**Number of pages**
arXiv.org	19
peer reviewed	12

Both versions of the paper seems to be largely identical, however, the preprint available on arXiv.org includes four Appendices that are not included in the peer reviewed version. This accounts for the missing 7 pages.

f) The number of times the article has been cited on Google Scholar and Web of Science is recorded below (Table 2).

Table 2: Number of times Zhang et al. has been cited.

Article version	**Number of Google Scholar citations**	**Number of Web of Science citations**	**Data(s) of access**
peer reviewed	779	29	26 Aug 2024

g) The place the top 10 most relevant citing papers were published are summarised in the table below (Table 3).

Table 3: Top 10 most relevant papers that cited Zhang et. al.

	Journals	**Refereed Conferences**	**Preprints and unrefereed conferences**	**Data(s) of access**
Title of publication of citing articles	Sensors Machine Learning: Science and Technology	3 × IEEE/CVF Conference on Computer Vision and Pattern Recognition 2 × NeurIPS Computer Vision ECCV 2020	2 × arXiv.org	26 August 2024
total count	2	6	2	

Table 3 shows that the majority of citing papers are from refereed conferences, with only a two papers being preprints. This does not line up with the difference in the number of citations between Google Scholar and Web of Science as it is expected that more preprints and unrefereed conferences would appear on the list of top 10 most relevant papers. This may be due to Google preferencing refereed conferences over unrefereed conferences and preprints when ordering by relevance.

3 References & Authors

a) The published article is a magazine article published to Science.

b) The number of citations as reported by Google Scholar and Web of Science can be seen below (Table 4), along with the number of authors on the preprint and official publication.

Table 4: Number of citations and authors on each paper

Article version	**Google Scholar citations**	**Web of Science citations**	**Data(s) of access**	**Number of authors on paper**
arXiv.org	n/a	n/a	n/a	25
Journal	36	2	29 August 2024	25

c) Google Scholar merges multiple versions of the same paper and sums the citations of the each version to create a single citation count [7].

d) For this question, I assume the papers have more than 21 authors.

Table 5: Number of authors to include in in-text citations, when there are a large number of authors.

	IEEE reference style	**APA reference style**
First in-text citation	Authors do not need to be specified. Up to three authors can be specified, followed by "et al." for a reference cited in text [8].	The last name of the primary author is provided, followed by "et al." [9].
Subsequent in-text citations	Authors do not need to be specified. Up to three authors can be specified, followed by "et al." for a reference cited in text [8].	The last name of the primary author is provided, followed by "et al." [9].
Format in reference list	All authors must be listed, up to six authors. If there are more than six authors, the first author is followed by "et al." [8].	The names of the first 19 authors are specified, followed by an ellipsis, followed by the name of the last author [9].

e) This is the in-text citation for the preprint version [10].

This is the in-text citation for the online publication version [11].

f) The corresponding author and first three authors are recorded below (Table 6), with their current affiliations and academic or professional position.

Table 6: Authors of the published paper.

Author order	**Name**	**Affiliation**	**Academic or professional position**
Corresponding author	Jan Brauner	University of Oxford [12]	PhD Student [12]
Author #1	Yoshua Bengio	Mila Quebec Artificial Intelligence Institute, Universit de Montrea [13]	Professor [13]
Author #2	Geoffrey Hinton	The University of Toronto [14]	Professor Emeritus [14]
Author #3	Andrew Yao	Tsinghua University [15]	Dean, Professor [15]

g) There are several reasons why a paper may have a large number of authors, including increasing the credibility of the publication, the publication resulting from a collective brainstorm, or increasing the reputation of the authors. Including authors that hold prestigious positions within a field (e.g. university deans and professors) can increase the credibility of the publication, justifying the inclusion of additional authors with strong reputations. Some publications are also a result of discussions between large groups of experts, such as policy forums, resulting in a large number of authors. Alternatively, some researchers convince their colleagues to include them in papers in order to inflate their research output, to increase their metrics like their h-index, with minimal effort [16].

In the case of this paper [11], Science labels the article as a policy forum. This means the article resulted from an analysis between a diverse group of field experts on "the policy implications of recent scientific results or studies" [17]. [11] was a discussion between numerous AI experts, with high standings within their field, on the implication of the rapid progress of AI on society.

4 Quality Metrics For Conferences & Conference Papers

a) A common conference-level metric for determining the quality of a conference is the rank provided on the ICORE Conference Portal. This rank is "determined by a mix of indicators, including citation rates, paper submission and acceptance rates, and the visibility and research track record of the key people hosting the conference and managing its technical program" [18].

Another common metric to consider is the h-index of the conference, which reflects the quality of the published papers. However, the h-index for IADIS is currently unavailable, likely because it is considered a less prominent conference. An alternative metric is Resurchify's star rating, which is based on the CORE ranking.

b) The table below (Table 7) shows the metrics and ANZSRC FoR codes for the provided conferences. Metric 1 is the ranking found on the ICORE Conference Portal. Metric 2 is the star rating found on Resurchify.

Table 7: ANZSRC Fields of Research (FoR) codes and metrics.

Conference	**Metric 1**	**Metric 2**	**ANZSRC FoR codes**
Association for Computational Linguistics	A* [19]	5/5 [20]	4602 [19]
IADIS International Conference Applied Computing	C [21]	2/5 [22]	4601 [21]

5 Conflicts of Interest

a) This would be a potential conflict of interest under Personal Relationships. Jim likely had a strong negative relationship with Elizabeth during their disagreement, which likely to affect their review of Elizabeth's application, making it a conflict of interest of a Personal Relationship type. Additionally, as Jim is informing the committee's decision but not directly involved, the UQ Conflict of Interest Policy [23] would define it as a potential conflict, instead of an actual conflict.

b) It is not a conflict of interest for the other members of the committee, such as the Chair, as they are not involved in Jim and Elizabeth's relationship. However, they may have other conflicts of interests with Jim or Elizabeth if there is an existing relationship between them.

c) According to the UQ Conflict of Interest Policy, Jim's "conflict of interest must be disclosed and managed transparently between relevant parties" [23]. As a result, it may not be required to disclose the conflict to the entire committee, however, Jim must disclose it to the Chair.

d) The Chair should ensure Jim formally registers his conflict of interest with the University and help implement a conflict of interest management plan. This may involve the Chair recruiting another member of the committee to review Elizabeth's application and restricting Jim's involvement in Elizabeth's potential interview.

e) An actual conflict of interest is when "a staff member has a competing interest or obligation ... that directly conflicts with the staff member's duties and responsibilities" [23]. In contrast, a perceived conflict of interest is when "it could reasonably be perceived, or give the appearance, that a competing interest or obligation ... could improperly influence the performance of a staff member's duties and responsibilities" [23]. Both types of conflicts of interest must be addressed by researchers and research institutions as they can negatively effect the credibility of the produced work.

6 Demonstrating Your Use of A Reference Manager

References

[1] A. Hofer, B. Caruso, and S. Brinn. “About arXiv - arXiv info.” arXiv.org. Accessed: Aug. 26, 2024. [Online]. Available: https://info.arxiv.org/about/index.html

[2] J. Stevens. “University Library: Grey literature: Conference proceedings.” Accessed: Aug. 26, 2024. [Online]. Available: https://library.nd.edu.au/instruction/greyliterature/conference

[3] P. Coulter. “What does ‘peer reviewed’ mean? - LibAnswers.” LibAnswers. Accessed: Aug. 29, 2024. [Online]. Available: https://apus.libanswers.com/faq/2154

[4] Web of Science. “Web of Science Core Collection: Document Type Descriptions.” Clarivate. Accessed: Aug. 29, 2024. [Online]. Available: https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-Core-Collection-Document-Type-Descriptions?language=en_US

[5] M. R. Zhang, J. Lucas, G. Hinton, and J. Ba, “Lookahead optimizer: k steps forward, 1 step back,” 2019, *arXiv*: 1907.08610.

[6] M. Zhang, J. Lucas, J. Ba, and G. E. Hinton, “Lookahead optimizer: k steps forward, 1 step back,” in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett, Eds., Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/90fd4f88f588ae64038134f1eeaa023f-Paper.pdf

[7] H. L. Bodlaender and M. van Kreveld, “Google Scholar makes it hard – the complexity of organizing one’s publications,” *Information Processing Letters*, vol. 115, no. 12, pp. 965–968, Dec. 2015, doi: 10.1016/j.ipl.2015.07.003.

[8] IEEE, “Reference Guide.” IEEE Publication Operations, Nov. 29, 2023. Accessed: Aug. 29, 2024. [Online]. Available: http://journals.ieeeauthorcenter.ieee.org/wp-content/uploads/sites/7/IEEE_Reference_Guide.pdf

[9] The University of Queensland. “Library Guides: APA 7th referencing style: Journal article.” Library Guides at University of Queensland Library. Accessed: Aug. 29, 2024. [Online]. Available: https://guides.library.uq.edu.au/referencing/apa7/journal-articles

[10] Y. Bengio *et al.*, “Managing extreme AI risks amid rapid progress,” Oct. 26, 2023, *arXiv*: 2310.17688v.

[11] Y. Bengio *et al.*, “Managing extreme AI risks amid rapid progress,” *Science*, vol. 384, no. 6698, pp. 842–845, May, 2024. Accessed: Aug. 29, 2024. [Online]. Available: https://www.science.org/doi/10.1126/science.adn0117

[12] J. Brauner. “Jan Brauner - PHD Student - University of Oxford | LinkedIn.” LinkedIn. Accessed: Aug. 29, 2024. [Online]. Available: https://uk.linkedin.com/in/jan-brauner-4679a3146

[13] Y. Bengio. “Yoshua Bengio.” Yoshua Bengio. Accessed: Aug. 29, 2024. [Online]. Available: https://yoshuabengio.org/

[14] University College. “Geoffery Hinton.” University College U of T. Accessed: Aug. 30, 2024. [Online]. Available: https://www.uc.utoronto.ca/staff-faculty-profile/geoffrey-hinton

[15] Institute for Interdisciplinary Information Sciences. “姚期智 / Andrew Chi-Chih Ya.” Institute for Interdisciplinary Information Sciences (IIIS). [Online]. Available: https://iiis.tsinghua.edu.cn/en/people/#jy1

[16] The Economist, “Why research papers have so many authors,” *The Economist*, Nov, 2016. Accessed: Aug. 30, 2024. [Online]. Available: https://www.economist.com/science-and-technology/2016/11/24/why-research-papers-have-so-many-authors

[17] B. Hansin, “Instructions for Reviewers of Policy Forums-1630080681.” American Association for the Advancement of Science, 2019. Accessed: Aug. 30, 2024. [Online]. Available: https://www.science.org/pb-assets/PDF/Information%20for%20reviewers/Instructions%20for%20Reviewers%20of%20Policy%20Forums-1630080681.pdf

[18] M. Winikoff, A. Fekete, S. MacDonell, L. Padgham, and S. Zollmann. “CORE Rankings Portal.” core.edu.au. Accessed: Aug. 30, 2024. [Online]. Available: https://www.core.edu.au/conference-portal

[19] ICORE. “Association for Computational Linguistics.” ICORE Conference Portal. Accessed: Aug. 30, 2024. [Online]. Available: https://portal.core.edu.au/conf-ranks/196/

[20] Resurchify. “ACL : Association of Computational Linguistics Ranking and Rating | Resurchify.” Resurchify. Accessed: Aug. 30, 2024. [Online]. Available: https://www.resurchify.com/conference_ranking_details.php?id=252

[21] ICORE. “IADIS International Conference Applied Computing.” ICORE Conference Portal. Accessed: Aug. 30, 2024. [Online]. Available: https://portal.core.edu.au/conf-ranks/2036/

[22] Resurchify. “IADIS AC : IADIS International Conference Applied Computing Ranking and Rating | Resurchify.” Resurchify. Accessed: Aug. 30, 2024. [Online]. Available: https://www.resurchify.com/conference_ranking_details.php?id=956

[23] P. Vaughan. “Conflict of Interest Policy.” Policy and Procedure Library. Accessed: Aug. 29, 2024. [Online]. Available: https://policies.uq.edu.au/document/view-current.php?id=68

A Bibtex File

```
@misc{resurchify_iadis_2022,
        title = {{IADIS} {AC} : {IADIS} {International} {Conference}
            {Applied} {Computing} {Ranking} and {Rating} {\textbar}
           {Resurchify}},
        url = {https://www.resurchify.com/conference_ranking_details
           .php?id=956},
        urldate = {2024-08-30},
        journal = {Resurchify},
        author = {{Resurchify}},
        year = {2022},
}

@misc{resurchify_acl_2022,
        title = {{ACL} : {Association} of {Computational} {
           Linguistics} {Ranking} and {Rating} {\textbar} {
           Resurchify}},
        url = {https://www.resurchify.com/conference_ranking_details
           .php?id=252},
        urldate = {2024-08-30},
        journal = {Resurchify},
        author = {{Resurchify}},
        year = {2022},
}

@misc{winikoff_coreeduau_2024,
        title = {core.edu.au - {CORE} {Rankings} {Portal}},
        url = {https://www.core.edu.au/conference-portal},
        abstract = {ACCESS THE CORE CONFERENCE DB HERE
ACCESS THE CORE JOURNAL DB HERE  DISCONTINUED FROM MARCH 2022
EXPLORE WHERE PEOPLE PUBLISH  HERE
USING THE CONFERENCE AND JOURNALS DATABASES},
        language = {en-US},
        urldate = {2024-08-30},
        journal = {CORE Rankings Portal},
        author = {Winikoff, Michael and Fekete, Alan and MacDonell,
           Stephen and Padgham, Lin and Zollmann, Stefanie},
```

```
        month = mar,
        year = {2024},
}

@misc{icore_iadis_2023,
        title = {{IADIS} {International} {Conference} {Applied} {
           Computing}},
        url = {https://portal.core.edu.au/conf-ranks/2036/},
        urldate = {2024-08-30},
        journal = {ICORE Conference Portal},
        author = {{ICORE}},
        year = {2023},
}

@misc{icore_association_2023,
        title = {Association for {Computational} {Linguistics}},
        url = {https://portal.core.edu.au/conf-ranks/196/},
        urldate = {2024-08-30},
        journal = {ICORE Conference Portal},
        author = {{ICORE}},
        year = {2023},
}

@misc{hansin_instructions_2019,
        title = {Instructions for {Reviewers} of {Policy} {Forums
           }-1630080681},
        url = {https://www.science.org/pb-assets/PDF/Information%20
           for%20reviewers/Instructions%20for%20Reviewers%20of%20
           Policy%20Forums-1630080681.pdf},
        urldate = {2024-08-30},
        publisher = {American Association for the Advancement of
           Science},
        author = {Hansin, Brooks},
        year = {2019},
}

@article{the_economist_why_2016,
        title = {Why research papers have so many authors},
        issn = {0013-0613},
        url = {https://www.economist.com/science-and-technology
```

```
        /2016/11/24/why-research-papers-have-so-many-authors},
    urldate = {2024-08-30},
    journal = {The Economist},
    author = {{The Economist}},
    month = nov,
    year = {2016},
}

@misc{institute_for_interdisciplinary_information_sciences__2024,
    title = {          / {Andrew} {Chi}-{Chih} {Ya}},
    url = {https://iiis.tsinghua.edu.cn/en/people/#jy1},
    journal = {IIIS, Tsinghua University},
    author = {{Institute for Interdisciplinary Information
        Sciences}},
    year = {2024},
}

@misc{university_college_geoffery_2024,
    title = {Geoffery {Hinton}},
    url = {https://www.uc.utoronto.ca/staff-faculty-profile/
        geoffrey-hinton},
    urldate = {2024-08-29},
    journal = {University College},
    author = {{University College}},
    year = {2024},
}

@misc{bengio_yoshua_2024,
    title = {Yoshua {Bengio}},
    url = {https://yoshuabengio.org/},
    abstract = {Yoshua Bengio est reconnu comme l un des plus
        grands experts mondiaux en mati re d intelligence
        artificielle et comme un pionnier de l apprentissage
        profond.},
    language = {en-CA},
    urldate = {2024-08-29},
    journal = {Yoshua Bengio},
    author = {Bengio, Yoshua},
    month = jul,
    year = {2024},
```

```
}

@misc{brauner_jan_2024,
        title = {Jan {Brauner} - {PHD} {Student} - {University} of {
           Oxford} {\textbar} {LinkedIn}},
        url = {https://uk.linkedin.com/in/jan-brauner-4679a3146},
        abstract = {Machine learning researcher, medical doctor
           Very ragtag background: I used to do medicine, and now I
           \&\#39;m a machine learning researcher.\&lt;br\&gt;\&lt;
           br\&gt;Research career highlights:\&lt;br\&gt;      \&
           amp;gt;25 peer-reviewed publications, \&amp;gt;2,500
           citations.\&lt;br\&gt;          Published in high-
           profile outlets, e.g. Science, Nature Communications,
           PNAS, ICML, ICLR, NeurIPS (spotlight).\&lt;br\&gt;
           My COVID work included the 7th most-discussed paper in
           the history of the journal Science (at the time of
           publication), my work on data selection included the most
            popular arXiv paper of the day.\&lt;br\&gt;
           Attracted \&amp;gt; 300,000 in third-party funding.\&lt
           ;br\&gt;      Work cited in federal bills and presented
           at institutions, e.g. Africa CDC, OECD Global Science
           Forum, UK Cabinet Office.\&lt;br\&gt;          Advised
           top-level politicians on AI, e.g. a German Secretary of
           State. \&lt;br\&gt;     Invited talks at OpenAI, Meta AI
           Research, various universities.\&lt;br\&gt;    Media
           appearances and work featured in major newspapers, radio
           and TV stations, e.g. Forbes, Guardian, Vox, TIME.\&lt;br
           \&gt;          Received award for largest social impact
           among researchers of any career stage at Oxford s STEM
           departments.\&lt;br\&gt;\&lt;br\&gt;\&lt;br\&gt;
           Publications: https://scholar.google.de/citations?user
           \&\#61;tNZUnjcAAAAJ\&amp;hl\&\#61;de\&amp;oi\&\#61;sra
            Experience: University of Oxford    Education:
           University of Oxford    Location: Oxford    477
           connections on LinkedIn. View Jan Brauner s profile on
           LinkedIn, a professional community of 1 billion members
           .},
        language = {en},
        urldate = {2024-08-29},
        journal = {LinkedIn},
```

```
        author = {Brauner, Jan},
        year = {2024},
}

@misc{the_university_of_queensland_library_2024,
        title = {Library {Guides}: {APA} 7th referencing style: {
           Journal} article},
        copyright = {Copyright University of Queensland Library
           2024},
        shorttitle = {Library {Guides}},
        url = {https://guides.library.uq.edu.au/referencing/apa7/
           journal-articles},
        abstract = {This is a guide to using the APA7 referencing
           style from the American Psychological Association. It is
           based on the Publication Manual of the American
           Psychological Association.},
        language = {en},
        urldate = {2024-08-29},
        journal = {The University of Queensland},
        author = {{The University of Queensland}},
        month = aug,
        year = {2024},
}

@misc{ieee_reference_2023,
        title = {Reference {Guide}},
        copyright = {IEEE},
        url = {http://journals.ieeeauthorcenter.ieee.org/wp-content/
           uploads/sites/7/IEEE_Reference_Guide.pdf},
        urldate = {2024-08-29},
        publisher = {IEEE Publication Operations},
        author = {{IEEE}},
        month = nov,
        year = {2023},
}

@article{bodlaender_google_2015,
        title = {Google {Scholar} makes it hard     the complexity
           of organizing one's publications},
        volume = {115},
```

```
        issn = {0020-0190},
        url = {https://www.sciencedirect.com/science/article/pii/
           S0020019015001167},
        doi = {10.1016/j.ipl.2015.07.003},
        abstract = {With Google Scholar, scientists can maintain
           their publications on personal profile pages, while the
           citations to these works are automatically collected and
           counted. Maintenance of publications is done manually by
           the researcher herself, and involves deleting erroneous
           ones, merging ones that are the same but which were not
           recognized as the same, adding forgotten co-authors, and
           correcting titles of papers and venues. The publications
           are presented on pages with 20 or 100 papers in the web
           page interface from 2012 2014. (Since mid 2014, Google
           Scholar's profile pages allow any number of papers on a
           single page.) The interface does not allow a scientist to
            merge two versions of a paper if they appear on
           different pages. This not only implies that a scientist
           who wants to merge certain subsets of publications will
           sometimes be unable to do so, but also, we show in this
           note that the decision problem to determine if it is
           possible to merge given subsets of papers is NP-complete
           .},
        number = {12},
        urldate = {2024-08-29},
        journal = {Information Processing Letters},
        author = {Bodlaender, Hans L. and van Kreveld, Marc},
        month = dec,
        year = {2015},
        keywords = {3-partition, Computational complexity, Google
           Scholar, NP-completeness, Reduction},
        pages = {965--968},
}

@article{bengio_managing_2024,
        title = {Managing extreme {AI} risks amid rapid progress},
        volume = {384},
        issn = {0036-8075, 1095-9203},
        url = {https://www.science.org/doi/10.1126/science.adn0117},
        abstract = {Preparation requires technical research and
```

```
    development, as well as adaptive, proactive governance
    ,

        Artificial intelligence (AI) is progressing rapidly,
           and companies are shifting their focus to
           developing generalist AI systems that can
           autonomously act and pursue goals. Increases in
           capabilities and autonomy may soon massively
           amplify AI s impact, with risks that include
           large-scale social harms, malicious uses, and an
           irreversible loss of human control over autonomous
           AI systems. Although researchers have warned of
           extreme risks from AI (

             1

        ), there is a lack of consensus about how to manage
           them. Society s response, despite promising first
            steps, is incommensurate with the possibility of
           rapid, transformative progress that is expected by
           many experts. AI safety research is lagging.
           Present governance initiatives lack the mechanisms
           and institutions to prevent misuse and recklessness
            and barely address autonomous systems. Drawing on
           lessons learned from other safety-critical
           technologies, we outline a comprehensive plan that
           combines technical research and development (R\&D)
           with proactive, adaptive governance mechanisms for
           a more commensurate preparation.},
language = {en},
number = {6698},
urldate = {2024-08-29},
journal = {Science},
author = {Bengio, Yoshua and Hinton, Geoffrey and Yao,
    Andrew and Song, Dawn and Abbeel, Pieter and Darrell,
    Trevor and Harari, Yuval Noah and Zhang, Ya-Qin and Xue,
    Lan and Shalev-Shwartz, Shai and Hadfield, Gillian and
    Clune, Jeff and Maharaj, Tegan and Hutter, Frank and
    Baydin, A t l m G n e  and McIlraith, Sheila and Gao,
    Qiqi and Acharya, Ashwin and Krueger, David and Dragan,
```

```
      Anca and Torr, Philip and Russell, Stuart and Kahneman,
      Daniel and Brauner, Jan and Mindermann, S ren},
    month = may,
    year = {2024},
    pages = {842--845},
}

@misc{bengio_managing_2023,
    title = {Managing extreme {AI} risks amid rapid progress},
    url = {http://arxiv.org/abs/2310.17688},
    abstract = {Artificial Intelligence (AI) is progressing
      rapidly, and companies are shifting their focus to
      developing generalist AI systems that can autonomously
      act and pursue goals. Increases in capabilities and
      autonomy may soon massively amplify AI's impact, with
      risks that include large-scale social harms, malicious
      uses, and an irreversible loss of human control over
      autonomous AI systems. Although researchers have warned
      of extreme risks from AI, there is a lack of consensus
      about how exactly such risks arise, and how to manage
      them. Society's response, despite promising first steps,
      is incommensurate with the possibility of rapid,
      transformative progress that is expected by many experts.
       AI safety research is lagging. Present governance
      initiatives lack the mechanisms and institutions to
      prevent misuse and recklessness, and barely address
      autonomous systems. In this short consensus paper, we
      describe extreme risks from upcoming, advanced AI systems
      . Drawing on lessons learned from other safety-critical
      technologies, we then outline a comprehensive plan
      combining technical research and development with
      proactive, adaptive governance mechanisms for a more
      commensurate preparation.},
    language = {en},
    urldate = {2024-08-29},
    publisher = {arXiv},
    author = {Bengio, Yoshua and Hinton, Geoffrey and Yao,
      Andrew and Song, Dawn and Abbeel, Pieter and Darrell,
      Trevor and Harari, Yuval Noah and Zhang, Ya-Qin and Xue,
      Lan and Shalev-Shwartz, Shai and Hadfield, Gillian and
```

```
        Clune, Jeff and Maharaj, Tegan and Hutter, Frank and
        Baydin, Atlm Gne and McIlraith, Sheila and Gao,
        Qiqi and Acharya, Ashwin and Krueger, David and Dragan,
        Anca and Torr, Philip and Russell, Stuart and Kahneman,
        Daniel and Brauner, Jan and Mindermann, Sren},
    month = oct,
    year = {2023},
    note = {arXiv:2310.17688 [cs]},
    keywords = {Computer Science - Artificial Intelligence,
        Computer Science - Computation and Language, Computer
        Science - Computers and Society, Computer Science -
        Machine Learning},
}

@misc{web_of_science_web_2023,
    title = {Web of {Science} {Core} {Collection}: {Document} {
        Type} {Descriptions}},
    url = {https://support.clarivate.com/
        ScientificandAcademicResearch/s/article/Web-of-Science-
        Core-Collection-Document-Type-Descriptions?language=en_US
        },
    urldate = {2024-08-29},
    journal = {Clarivate},
    author = {{Web of Science}},
    month = feb,
    year = {2023},
}

@misc{coulter_what_2022,
    type = {Forum},
    title = {What does "peer reviewed" mean? - {LibAnswers}},
    url = {https://apus.libanswers.com/faq/2154},
    urldate = {2024-08-29},
    journal = {LibAnswers},
    author = {Coulter, Priscilla},
    month = jul,
    year = {2022},
}

@misc{vaughan_conflict_2022,
```

```
        title = {Conflict of {Interest} {Policy}},
        copyright = {The University of Queensland},
        url = {https://policies.uq.edu.au/document/view-current.php?
           id=68},
        urldate = {2024-08-29},
        journal = {Policy and Procedure Library},
        author = {Vaughan, Phil},
        month = oct,
        year = {2022},
}

@inproceedings{zhang_lookahead_2019,
        title = {Lookahead optimizer: k steps forward, 1 step back},
        volume = {32},
        url = {https://proceedings.neurips.cc/paper_files/paper
           /2019/file/90fd4f88f588ae64038134f1eeaa023f-Paper.pdf},
        booktitle = {Advances in {Neural} {Information} {Processing}
            {Systems}},
        publisher = {Curran Associates, Inc.},
        author = {Zhang, Michael and Lucas, James and Ba, Jimmy and
           Hinton, Geoffrey E},
        editor = {Wallach, H. and Larochelle, H. and Beygelzimer, A.
            and Alch -Buc, F. d' and Fox, E. and Garnett, R.},
        year = {2019},
}

@misc{zhang_lookahead_2019-1,
        title = {Lookahead optimizer: k steps forward, 1 step back},
        shorttitle = {Lookahead {Optimizer}},
        abstract = {The vast majority of successful deep neural
           networks are trained using variants of stochastic
           gradient descent (SGD) algorithms. Recent attempts to
           improve SGD can be broadly categorized into two
           approaches: (1) adaptive learning rate schemes, such as
           AdaGrad and Adam, and (2) accelerated schemes, such as
           heavy-ball and Nesterov momentum. In this paper, we
           propose a new optimization algorithm, Lookahead, that is
           orthogonal to these previous approaches and iteratively
           updates two sets of weights. Intuitively, the algorithm
           chooses a search direction by looking ahead at the
```

```
        sequence of  fast weights" generated by another
        optimizer. We show that Lookahead improves the learning
        stability and lowers the variance of its inner optimizer
        with negligible computation and memory cost. We
        empirically demonstrate Lookahead can signi cantly
        improve the performance of SGD and Adam, even with their
        default hyperparameter settings on ImageNet, CIFAR10/100,
         neural machine translation, and Penn Treebank.},
      language = {en},
      publisher = {arXiv},
      author = {Zhang, Michael R. and Lucas, James and Hinton,
        Geoffrey and Ba, Jimmy},
      year = {2019},
      note = {arXiv:1907.08610 [cs, stat]},
      keywords = {Computer Science - Machine Learning, Computer
        Science - Neural and Evolutionary Computing, Statistics -
         Machine Learning},
}

@misc{stevens_university_2024,
      title = {University {Library}: {Grey} literature: {
        Conference} proceedings},
      copyright = {Copyright The University of Notre Dame
        Australia 2024},
      shorttitle = {University {Library}},
      url = {https://library.nd.edu.au/instruction/greyliterature/
        conference},
      abstract = {University Library: Grey literature: Conference
        proceedings},
      language = {en},
      urldate = {2024-08-26},
      author = {Stevens, Jackie},
      month = aug,
      year = {2024},
}

@misc{hofer_about_2024,
      title = {About {arXiv} - {arXiv} info},
      url = {https://info.arxiv.org/about/index.html},
      urldate = {2024-08-26},
```

```
        journal = {arXiv.org},
        author = {Hofer, Alison and Caruso, Brian and Brinn, Shamsi
           },
        month = may,
        year = {2024},
}

@article{soden_forces_1979,
        title = {Forces applied to a bicycle during normal cycling},
        volume = {12},
        doi = {10.1016/0021-9290(79)90041-1},
        abstract = {Forces that the rider applies to the pedals,
           saddle and handlebars during speeding, hill climbing and
           starting are estimated from cine film records using
           elementary mechanics. The results are compared with force
            measurements obtained from an instrumented pedal. Pedal
           forces of up to three times bodyweight were recorded
           during starting. Handlebar loads were always
           significantly large.},
        journal = {Journal of biomechanics},
        author = {Soden, P.D. and Adeyefa, Bodunrin},
        month = feb,
        year = {1979},
        pages = {527--41},
}

@misc{noauthor_how_2017,
        title = {How {Much} {Does} the {Average} {Man} {Weigh}?},
        url = {https://www.healthline.com/health/mens-health/average
           -weight-for-men},
        abstract = {Discover how much the average man weighs. Also
           find out how to use BMI and other tools to determine if
           you re at a healthy weight.},
        language = {en},
        urldate = {2024-08-23},
        journal = {Healthline},
        month = aug,
        year = {2017},
}
```

```
@misc{noauthor_how_nodate,
        title = {How to {Choose} the {Right} {Bike} {Pedals}},
        url = {https://www.cyclesuk.com/blog/how-to-choose-the-right
          -bike-pedals/},
        urldate = {2024-08-23},
}

@misc{noauthor_prowheel_nodate,
        title = {{PROWHEEL} {Ebike} {Crank} {Arm} {Available} in {
          Square} {Hole},{152MM}/{160MM}/{165MM}/{170MM}/{175MM} {
          Black} {Bike} {Crank} {Arm} for {Electric} {Bike},{Mid} {
          Mounted} {Motor},{Torque}(1 {Pair}) : {Amazon}.co.uk: {
          Sports} \& {Outdoors}},
        url = {https://www.amazon.co.uk/Prowheel-Bicycle-Available-
          Electric-Mountain/dp/B0B9KZPWZ1?th=1},
        urldate = {2024-08-23},
}

@misc{sperickson91_bicycle_2014,
        title = {Bicycle {Crankset} {Design} and {Analysis}},
        url = {https://setherickson.wordpress.com/2014/07/30/bicycle
          -crankset-design-analysis-2/},
        abstract = {In this project one had to design a bicycle
          crankset which consisted of right and left crank arms, a
          pulley, two pedals, and a lower bracket spindle. The
          design goal was to minimize the overall mas },
        language = {en},
        urldate = {2024-08-23},
        journal = {SETH ERICKSON},
        author = {{SPERICKSON91}},
        month = jul,
        year = {2014},
}

@misc{li_bike_2020,
        title = {Bike {Crank} {Arm} {Ultimate} {Guide} ({Right} {
          Length}, {Design} etc)},
        url = {https://www.firstcomponents.com/bike-crank-arm/},
        language = {en-US},
        urldate = {2024-08-22},
```

```
        author = {Li, Johnson},
        month = apr,
        year = {2020},
        note = {Section: Crankset},
}

@incollection{schaeffer_search_2000,
        title = {Search ideas in {Chinook}},
        booktitle = {Games in {AI} {Research}},
        publisher = {Department of Computing Science, University of
           Alberta},
        author = {Schaeffer, Jonathan},
        editor = {van den Herik, Jaap and Iida, Hiroyuki},
        year = {2000},
        pages = {19--30},
}

@inproceedings{bursztyn_when_2023,
        series = {Working {Paper} {Series}},
        title = {When {Product} {Markets} {Become} {Collective} {
           Traps}: {The} {Case} of {Social} {Media}},
        shorttitle = {When {Product} {Markets} {Become} {Collective}
            {Traps}},
        url = {https://www.nber.org/papers/w31771},
        doi = {10.3386/w31771},
        abstract = {Individuals might experience negative utility
           from not consuming a popular product. With such
           externalities to non-users, standard consumer surplus
           measures, which take aggregate consumption as given, fail
            to appropriately capture consumer welfare. We propose an
            approach to account for these externalities and apply it
            to estimate consumer welfare from two social media
           platforms: TikTok and Instagram. Incentivized experiments
            with college students indicate positive welfare based on
            the standard measure, but negative welfare when
           accounting for these non-user externalities. Our findings
            highlight the existence of product market traps, where
           large shares of active users prefer each platform not to
           exist.},
        urldate = {2024-08-08},
```

```
        publisher = {National Bureau of Economic Research},
        author = {Bursztyn, Leonardo and Handel, Benjamin R. and
           Jimenez, Rafael and Roth, Christopher},
        month = oct,
        year = {2023},
        doi = {10.3386/w31771},
}

@article{martin_would_2024,
        title = {Would you pay to quit {TikTok} and {Instagram}? {
           You}  d  be surprised how many would},
        shorttitle = {Would you pay to quit {TikTok} and {Instagram
           }?},
        url = {http://theconversation.com/would-you-pay-to-quit-
           tiktok-and-instagram-youd-be-surprised-how-many-would
           -235180},
        abstract = {Even though social media is free to use,
           research found many US university students would pay to
           quit it     especially TikTok     if they could beat
           their fear of missing out.},
        language = {en-AU},
        urldate = {2024-08-08},
        journal = {The Conversation},
        author = {Martin, Peter},
        month = jul,
        year = {2024},
}

@phdthesis{montalvo_aftereffects_1976,
        type = {phd},
        title = {Aftereffects, adaptation, and plasticity: a neural
           model for tunable feature space.},
        shorttitle = {Aftereffects, adaptation, and plasticity},
        url = {https://dl.acm.org/doi/abs/10.5555/908075},
        school = {University of Massachusetts Amherst},
        author = {Montalvo, Fanya S.},
        year = {1976},
        note = {AAI7706432},
}
```

```
@article{shrestha_review_2019,
        title = {Review of {Deep} {Learning} {Algorithms} and {
           Architectures}},
        volume = {7},
        url = {https://ieeexplore.ieee.org/abstract/document
           /8694781},
        doi = {10.1109/ACCESS.2019.2912200},
        journal = {IEEE Access},
        author = {Shrestha, Ajay and Mahmood, Ausif},
        year = {2019},
        keywords = {Computer architecture, Deep learning, Feature
           extraction, Feedforward neural networks, Machine learning
            algorithm, Recurrent neural networks, Training,
           artificial intelligence, backpropagation, convolution
           neural network, deep neural network architectures,
           optimization, supervised and unsupervised learning},
        pages = {53040--53065},
}

@article{cao_real-time_2015,
        title = {Real-time high-fidelity facial performance capture
           },
        volume = {34},
        issn = {0730-0301},
        url = {https://dl.acm.org/doi/10.1145/2766943},
        doi = {10.1145/2766943},
        abstract = {We present the first real-time high-fidelity
           facial capture method. The core idea is to enhance a
           global real-time face tracker, which provides a low-
           resolution face mesh, with local regressors that add in
           medium-scale details, such as expression wrinkles. Our
           main observation is that although wrinkles appear in
           different scales and at different locations on the face,
           they are locally very self-similar and their visual
           appearance is a direct consequence of their local shape.
           We therefore train local regressors from high-resolution
           capture data in order to predict the local geometry from
           local appearance at runtime. We propose an automatic way
           to detect and align the local patches required to train
           the regressors and run them efficiently in real-time. Our
```

```
        formulation is particularly designed to enhance the low-
        resolution global tracker with exactly the missing
        expression frequencies, avoiding superimposing spatial
        frequencies in the result. Our system is generic and can
        be applied to any real-time tracker that uses a global
        prior, e.g. blend-shapes. Once trained, our online
        capture approach can be applied to any new user without
        additional training, resulting in high-fidelity facial
        performance reconstruction with person-specific wrinkle
        details from a monocular video camera in real-time.},
    number = {4},
    urldate = {2024-08-08},
    journal = {ACM Trans. Graph.},
    author = {Cao, Chen and Bradley, Derek and Zhou, Kun and
        Beeler, Thabo},
    month = jul,
    year = {2015},
    pages = {46:1--46:9},
}
```

B LaTeX Generated Reference List

References

[1] A. Hofer, B. Caruso, and S. Brinn, "About arXiv - arXiv info," May 2024. [Online]. Available: https://info.arxiv.org/about/index.html

[2] J. Stevens, "University Library: Grey literature: Conference proceedings," Aug. 2024. [Online]. Available: https://library.nd.edu.au/instruction/greyliterature/conference

[3] P. Coulter, "What does "peer reviewed" mean? - LibAnswers," Jul. 2022. [Online]. Available: https://apus.libanswers.com/faq/2154

[4] Web of Science, "Web of Science Core Collection: Document Type Descriptions," Feb. 2023. [Online]. Available: https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-Core-Collection-Document-Type-Descriptions?language=en_US

[5] M. R. Zhang, J. Lucas, G. Hinton, and J. Ba, "Lookahead optimizer: k steps forward, 1 step back," 2019, arXiv:1907.08610 [cs, stat].

[6] M. Zhang, J. Lucas, J. Ba, and G. E. Hinton, "Lookahead optimizer: k steps forward, 1 step back," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d. Alch-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/90fd4f88f588ae64038134f1eeaa023f-Paper.pdf

[7] H. L. Bodlaender and M. van Kreveld, "Google Scholar makes it hard the complexity of organizing one's publications," *Information Processing Letters*, vol. 115, no. 12, pp. 965–968, Dec. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020019015001167

[8] IEEE, "Reference Guide," Nov. 2023. [Online]. Available: http://journals.ieeeauthorcenter.ieee.org/wp-content/uploads/sites/7/IEEE_Reference_Guide.pdf

[9] The University of Queensland, "Library Guides: APA 7th referencing style: Journal article," Aug. 2024. [Online]. Available: https://guides.library.uq.edu.au/referencing/apa7/journal-articles

[10] Y. Bengio, G. Hinton, A. Yao, D. Song, P. Abbeel, T. Darrell, Y. N. Harari, Y.-Q. Zhang, L. Xue, S. Shalev-Shwartz, G. Hadfield, J. Clune, T. Maharaj, F. Hutter, A. G. Baydin, S. McIlraith, Q. Gao, A. Acharya, D. Krueger, A. Dragan, P. Torr, S. Russell, D. Kahneman, J. Brauner, and S. Mindermann, "Managing extreme AI risks amid rapid progress," Oct. 2023, arXiv:2310.17688 [cs]. [Online]. Available: http://arxiv.org/abs/2310.17688

[11] ——, "Managing extreme AI risks amid rapid progress," *Science*, vol. 384, no. 6698, pp. 842–845, May 2024. [Online]. Available: https://www.science.org/doi/10.1126/science.adn0117

[12] J. Brauner, "Jan Brauner - PHD Student - University of Oxford | LinkedIn," 2024. [Online]. Available: https://uk.linkedin.com/in/jan-brauner-4679a3146

[13] Y. Bengio, "Yoshua Bengio," Jul. 2024. [Online]. Available: https://yoshuabengio.org/

[14] University College, "Geoffery Hinton," 2024. [Online]. Available: https://www.uc.utoronto.ca/staff-faculty-profile/geoffrey-hinton

[15] Institute for Interdisciplinary Information Sciences, " / Andrew Chi-Chih Ya," 2024. [Online]. Available: https://iiis.tsinghua.edu.cn/en/people/#jy1

[16] The Economist, "Why research papers have so many authors," *The Economist*, Nov. 2016. [Online]. Available: https://www.economist.com/science-and-technology/2016/11/24/why-research-papers-have-so-many-authors

[17] B. Hansin, "Instructions for Reviewers of Policy Forums-1630080681," 2019. [Online]. Available: https://www.science.org/pb-assets/PDF/Information%20for%20reviewers/Instructions%20for%20Reviewers%20of%20Policy%20Forums-1630080681.pdf

[18] M. Winikoff, A. Fekete, S. MacDonell, L. Padgham, and S. Zollmann, "core.edu.au - CORE Rankings Portal," Mar. 2024. [Online]. Available: https://www.core.edu.au/conference-portal

[19] ICORE, "Association for Computational Linguistics," 2023. [Online]. Available: https://portal.core.edu.au/conf-ranks/196/

[20] Resurchify, "ACL : Association of Computational Linguistics Ranking and Rating | Resurchify," 2022. [Online]. Available: https://www.resurchify.com/conference_ranking_details.php?id=252

[21] ICORE, "IADIS International Conference Applied Computing," 2023. [Online]. Available: https://portal.core.edu.au/conf-ranks/2036/

[22] Resurchify, "IADIS AC : IADIS International Conference Applied Computing Ranking and Rating | Resurchify," 2022. [Online]. Available: https://www.resurchify.com/conference_ranking_details.php?id=956

[23] P. Vaughan, "Conflict of Interest Policy," Oct. 2022. [Online]. Available: https://policies.uq.edu.au/document/view-current.php?id=68