



# The use of ChatGPT for identifying disruptive papers in science: a first exploration

Lutz Bornmann<sup>1</sup>  · Lingfei Wu<sup>2</sup> · Christoph Ettl<sup>1</sup>

Received: 10 September 2024 / Accepted: 27 September 2024 / Published online: 14 October 2024  
© The Author(s) 2024

## Abstract

ChatGPT has arrived in quantitative research evaluation. With the exploration in this Letter to the Editor, we would like to widen the spectrum of the possible use of ChatGPT in bibliometrics by applying it to identify disruptive papers. The identification of disruptive papers using publication and citation counts has become a popular topic in scientometrics. The disadvantage of the quantitative approach is its complexity in the computation. The use of ChatGPT might be an easy to use alternative.

**Keywords** Bibliometrics · ChatGPT · Disruption

ChatGPT has arrived in quantitative research evaluation. ChatGPT is “a generative text system that leverages machine learning to produce text that closely mimics human writing” (Torres-Salinas et al., 2024). Bornmann and Lepori (2024) propose to use ChatGPT for the task of finding comparable institutions (similar institutions) of a focal institution for conducting benchmark studies. de Winter (2024) explored the use of ChatGPT to predict citation counts, Mendeley readers, and social media engagement of research. Torres-Salinas et al. (2024) dealt with the application of ChatGPT “in bibliometric studies across five core research stages: preparation and consultation, data processing, data analysis, results interpretation, and scientific writing”. The results of these recently published exploration studies on the use of ChatGPT in bibliometrics are encouraging. We expect that ChatGPT will play an important role in quantitative research evaluation in the (near) future.

With the exploration in this Letter to the Editor, we would like to widen the spectrum of the possible use of ChatGPT in bibliometrics by applying it to identify disruptive papers.

---

✉ Lutz Bornmann  
bornmann@gv.mpg.de

Lingfei Wu  
liw105@pitt.edu

Christoph Ettl  
ettl@gv.mpg.de

<sup>1</sup> Science Policy and Strategy Department, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany

<sup>2</sup> School of Computing and Information, University of Pittsburgh, 135 N Bellefield Ave, Pittsburgh, PA 15213, United States

The identification of disruptive papers using publication and citation counts has become a popular topic, since Wu et al. (2019) introduced an index (D index) to measure the disruptiveness of papers. An overview of the research on the D index can be found in Leibel and Bornmann (2024). The index scores range from 1 (disruptive research) to  $-1$  (developing research). Since the D index of publications cannot be searched in the same way as citation counts in existing literature databases such as the Web of Science (Clarivate) or Scopus (Elsevier), we wondered whether there is a simple method based on artificial intelligence to identify disruptive publications. At the beginning of June 2024, we explored the use of ChatGPT by asking it the following question: “Which are the five most disruptive research papers between 1995 and 2000 in astrophysics?”

The answer was as follows: “Determining the five most disruptive research papers in astrophysics between 1995 and 2000 involves identifying works that significantly influenced the field, changed paradigms, or introduced groundbreaking ideas. Here are five such papers, widely recognized for their impact:

1. Perlmutter et al. (1999). “Measurements of  $\Omega$  and  $\Lambda$  from 42 High-Redshift Supernovae”. This paper provided key evidence for the accelerating expansion of the universe, leading to the concept of dark energy. It was foundational in earning the 2011 Nobel Prize in Physics for the discovery of the acceleration of the universe (<https://doi.org/10.1086/307221>).
2. Riess et al. (1998). “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant”. Another critical paper that, alongside Perlmutter’s work, confirmed the accelerating expansion of the universe and suggested the presence of dark energy (<https://doi.org/10.1086/300499>).
3. Mayor and Queloz (1995). “A Jupiter-mass companion to a solar-type star”. This paper reported the discovery of 51 Pegasi b, the first exoplanet discovered around a Sun-like star, which opened up a new field of extrasolar planet research (<https://doi.org/10.1038/378355a0>).
4. Genzel and Eckart (1996). “The Galactic Center Black Hole”. This research provided strong evidence for the existence of a supermassive black hole at the center of our galaxy, influencing the study of galactic nuclei and black holes (<https://doi.org/10.1098/rsta.1996.0003>).
5. Guth and Pi (1982) and Linde (1982, 1983) Inflation Theory Papers. While technically outside the given range, these foundational works on inflation theory were profoundly disruptive throughout the late ‘90s as they became widely accepted. The concept of cosmic inflation has been fundamental to our understanding of the early universe. Guth and Pi’s Paper: <https://doi.org/10.1103/PhysRevLett.49.1110>, Linde’s Papers: [https://doi.org/10.1016/0370-2693\(82\)91219-9](https://doi.org/10.1016/0370-2693(82)91219-9) and [https://doi.org/10.1016/0370-2693\(83\)90837-7](https://doi.org/10.1016/0370-2693(83)90837-7).

These papers had a significant and lasting impact on astrophysics, shaping the direction of research and expanding our understanding of the universe”.

Out of the seven papers mentioned by ChatGPT (with DOI), we found six in the WoS database. ChatGPT also included three papers in the reply that were out of the requested publication year defined in our prompt (but ChatGPT explained this procedure). Referring to the first and second papers mentioned by ChatGPT (Perlmutter et al., 1999; Riess et al., 1998), Saul Perlmutter was awarded one half, and Adam Guy Riess and another physicist Brian Paul Schmidt, each received a quarter of the 2011 Nobel Prize in physics for the

discovery of the universe's accelerating expansion. Both papers mentioned by ChatGPT are cited in the advanced information on the 2011 Nobel Prize in Physics (see [www.nobelprize.org](http://www.nobelprize.org)) as the two “breakthrough papers” in the field.

Michel Mayor and Didier Queloz were jointly awarded one half of the 2019 Nobel Prize in physics. Their paper (Mayor & Queloz, 1995) was also cited in the advanced information (see [www.nobelprize.org](http://www.nobelprize.org)) as the “breakthrough paper” on the first discovery of an exoplanet orbiting a solar-type star. The paper started a revolution of exoplanet research in astronomy. Referring to paper 4 by ChatGPT “Genzel and Eckart (1996)” the title ChatGPT provided is wrong. The given DOI belongs to a paper that is not related to the context, describing flow in a cascade of steam turbine blades (White et al., 1997). We found that both Reinhard Genzel and Andreas Eckart appear as co-authors in more than 100 papers, but there are only four papers in which the two were the only co-authors. One of them was published in 1996: Eckart and Genzel (1996). In line with the title ChatGPT provided, this paper reports evidence for a massive black hole at the centre of our Galaxy, and is consequently cited in the advanced information on the 2020 physics Nobel prize. One quarter of the Nobel prize was awarded to Reinhard Genzel. The analysis of the papers that come in fifth place in the ChatGPT list revealed that the authors Alan Guth, Andrei Linde, and Alexei Starobinsky have received the 2014 Kavli prize in astrophysics. Alan Guth's paper identified by ChatGPT (Guth & Pi, 1982) is his second most cited work. Linde's papers (Linde, 1982, 1983) represent his two most cited publications. The three scientists are widely recognized for introducing the concept of inflation in cosmology, a fundamental contribution to the Big Bang theory as the standard model of the universe.

In addition to face validity, the disruptive nature of the papers mentioned by ChatGPT is supported by our bibliometrics analysis. Table 1 includes the D index values and citation counts for the papers identified by ChatGPT. The D index values and citation counts

**Table 1** D index values and citation counts for the papers identified by ChatGPT as disruptive in astrophysics, based on the dataset shared by Lin et al. (2023)

Paper	D index value	Citation counts
Perlmutter et al. (1999) “Measurements of $\Omega$ and $\Lambda$ from 42 High-Redshift Supernovae”	–0.23 (bottom 13%)	9797 (top 1%)
Riess et al. (1998) “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant”	0.106 (top 4%)	10,093 (top 1%)
Mayor and Queloz (1995) “A Jupiter-mass companion to a solar-type star”	0.276 (top 2%)	1969 (top 1%)
Eckart and Genzel (1996) “Observations of stellar proper motions near the Galactic Centre”	0.003 (top 17%)	195 (top 2%)
Guth and Pi (1982) “Fluctuations in the new inflationary universe”	–	–
Linde (1982) “A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems”	0.180 (top 3%)	2892 (top 1%)
Linde (1983) “Chaotic inflation”	–	–

We exchanged “Genzel and Eckart (1996)”—mentioned by ChatGPT—with Eckart and Genzel (1996), since the mentioned paper seems to combine multiple papers, with Eckart and Genzel (1996) being the main source

are from a dataset shared by Lin et al. (2023). The bottom x% and top x% classifications in the table are based on the ranking position of a focal paper among the approximately 20 million papers in the dataset. Two disruptive papers are not covered in the dataset. As the results in Table 1 show, four papers have a positive and high D index value (top 5%). Only one paper, Perlmutter et al. (1999), has a negative D index value. But the negative value for a paper assessed by ChatGPT as disruptive seems to be understandable: Perlmutter et al. (1999) makes a great contribution by consolidating Riess et al. (1998) on the universe's accelerating expansion. The citation counts in Table 1 are all at least among the top 10% (four papers belong to the top 1% papers).

We noticed varying results by ChatGPT when repeating our request after June 2024. In order to explore the reasons for the varying results, we asked ChatGPT multiple times in the next weeks. Our explorations did not uncover the reasons but revealed changes in the sources ChatGPT uses for its responses. After our initial request in June 2024, ChatGPT has provided its used sources for the later replies. In the early replies, ChatGPT identified disruptive research using two sources: a Bing web search and the paper entitled "Astronomy's greatest hits: the 100 most cited papers in each year of the first decade of the twenty first century (2000–2009)" by Frogel (2010). ChatGPT gave the (wrong) impression that it included papers in the response to our requests that indicated Frogel (2010) as highly cited. We found that Frogel (2010) focused on publication years (from 2000 to 2009) that we were (mostly) not interested in (we asked for the years from 1995 to 2000). In later replies by ChatGPT, we observed not only an increase in the waiting times for our requests, but also that ChatGPT's responses changed significantly. For example, in the most recent requests, ChatGPT provided results from papers that have investigated the disruptiveness of research based on the D index. In another reply, ChatGPT explained (step by step) how certain datasets can be analyzed to identify disruptive papers.

Although we know that responses by generative text systems (such as ChatGPT) triggered by a prompt will always, in the zeroth approximation, return a very (or the most) probable sequence of characters and words in the selected context, we believe—based on our first explorations—that ChatGPT shows potential in identifying disruptive papers. Since the system is not perfect (as our results show), it should only be used by someone who can assess the quality of its replies. ChatGPT offers a much lower-cost alternative compared to using datasets including complex indicators, such as the D index, for empirical analyses. Beyond practical implications, our little experiment raises intriguing theoretical questions: How does ChatGPT select and integrate information across sources? Does it rely on paper contents, others' (scientists') evaluations, or citation relationships, similar to formal bibliometrics? These questions suggest directions for future research.

**Acknowledgements** Lutz Bornmann is member of the editorial board of *Scientometrics*. Lingfei Wu is grateful for support from the National Science Foundation grant SOS:DCI 2239418.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

Bornmann, L., & Lepori, B. (2024). The use of ChatGPT to find similar institutions for institutional benchmarking. *Scientometrics*. <https://doi.org/10.1007/s11192-024-05039-7>

de Winter, J. (2024). Can ChatGPT be used to predict citation counts, readership, and social media interaction? An exploration among 2222 scientific abstracts. *Scientometrics*, 129(4), 2469–2487. <https://doi.org/10.1007/s11192-024-04939-y>

Eckart, A., & Genzel, R. (1996). Observations of stellar proper motions near the Galactic Centre. *Nature*, 383(6599), 415–417. <https://doi.org/10.1038/383415a0>

Frogel, J. A. (2010). Astronomy's greatest hits: The 100 most cited papers in each year of the first decade of the 21st Century (2000–2009). Retrieved August 1, 2024, from <https://arxiv.org/abs/1005.5377>.

Guth, A. H., & Pi, S.-Y. (1982). Fluctuations in the new inflationary universe. *Physical Review Letters*, 49(15), 1110–1113. <https://doi.org/10.1103/PhysRevLett.49.1110>

Leibel, C., & Bornmann, L. (2024). What do we know about the disruption index in scientometrics? An overview of the literature. *Scientometrics*, 129, 601–639. <https://doi.org/10.1007/s11192-023-04873-5>

Lin, Y., Frey, C. B., & Wu, L. (2023). Remote collaboration fuses fewer breakthrough ideas. *Nature*, 623(7989), 987–991. <https://doi.org/10.1038/s41586-023-06767-1>

Linde, A. D. (1982). A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, 108(6), 389–393. [https://doi.org/10.1016/0370-2693\(82\)91219-9](https://doi.org/10.1016/0370-2693(82)91219-9)

Linde, A. D. (1983). Chaotic inflation. *Physics Letters B*, 129(3), 177–181. [https://doi.org/10.1016/0370-2693\(83\)90837-7](https://doi.org/10.1016/0370-2693(83)90837-7)

Mayor, M., & Queloz, D. (1995). A Jupiter-mass companion to a solar-type star. *Nature*, 378(6555), 355–359. <https://doi.org/10.1038/378355a0>

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., et al. (1999). Measurements of  $\Omega$  and  $\Lambda$  from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2), 565. <https://doi.org/10.1086/307221>

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., et al. (1998). Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3), 1009. <https://doi.org/10.1086/300499>

Torres-Salinas, D., Thelwall, M., & Arroyo-Machado, W. (2024). ChatGPT for bibliometrics: A comprehensive corpus of applications. <https://doi.org/10.5281/zenodo.11103551>

White, A. J., Young, J. B., & Walters, P. T. (1997). Experimental validation of condensing flow theory for a stationary cascade of steam turbine blades. *Philosophical Transactions of the Royal Society of London. Series a: Mathematical, Physical and Engineering Sciences*, 354(1704), 59–88. <https://doi.org/10.1098/rsta.1996.0003>

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, 566, 378–382. <https://doi.org/10.1038/s41586-019-0941-9>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.