



ILLUSTRATION: CHIARA VERGESI

SCIENCE'S GOLDEN OLDIES: THE DECADES-OLD RESEARCH PAPERS STILL HEAVILY CITED TODAY

An analysis for *Nature* reveals the studies that appear most in the reference lists of current publications.

By Richard Van Noorden

Researchers advance by standing on the shoulders of giants, to paraphrase Isaac Newton. So, which research giants are still getting cited frequently today?

One way of answering that is to determine which articles appear most often in the reference lists of today's research papers. *Nature* asked three bibliometricians who study patterns of references in scientific publications to dig into the data. They churned through tens of millions of references cited in all the papers published in 2023, the most complete year

WHICH ARTICLES ARE REFERENCED THE MOST?

The top-ten cited references across three databases* for all papers published in 2023.

Rank	Title	Number of citations (range)
1	Deep residual learning for image recognition (2016)	19,826–33,339
2	Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries (2021)	10,952–24,830
3	Attention is all you need (2017)	9,395–19,348
4	Generalized gradient approximation made simple (1996)	14,338–17,540
5	Using thematic analysis in psychology (2006)	10,660–17,347
6	The PRISMA 2020 statement: an updated guideline for reporting systematic reviews (2021)	13,115–13,443
7	Analysis of relative gene expression data using real-time quantitative PCR and the 2 ^{-ΔΔC_T} method (2001)	11,851–13,082
8	Random forests (2001)	6,266–12,294
9	Efficient iterative schemes for <i>ab initio</i> total-energy calculations using a plane-wave basis set (1996)	10,063–10,789
10	Long short-term memory (1997)	2,394–12,355

*Web of Science, OpenAlex and Scopus. Ordered according to median of ranks in the three databases.

For methodology and full data, see ref. 2 and Supplementary information (go.nature.com/4j5xtfh).

available in research repositories at the time.

The researchers found that popular twenty-first-century papers on topics such as artificial intelligence (AI), scientific software and methods to improve the quality of research dominate today's reference lists. But some studies published before 2000 are still heavily acknowledged even now (see 'Which articles are referenced the most?').

Studying the cited references of a large collection of papers provides a foundation for finding the intellectual roots of a field, says Robin Haunschild, a scientometrician at the Max Planck Institute for Solid State Research in Stuttgart, Germany.

"This reveals the really important papers for a community," adds Lutz Bornmann, a sociologist of science at the Max Planck Society in Munich, Germany. Bornmann and Haunschild, together with Andreas Thor, a computer scientist at the Leipzig University of Applied Sciences, Germany, conducted the technically challenging analysis for *Nature*. They relied on a modified version of software that Bornmann and Thor, with others, introduced in 2016 to help researchers to explore cited references¹.

To compile a list of the most-cited references in publications today, *Nature* took the median rankings across three databases (see the scientometricians' methodology² and Supplementary information).

AI influencers

The analysis reveals that work on AI published in the past decade was referenced the most. *Nature* found a similar pattern in a related exploration of the most-cited papers published in the twenty-first century (see page 588).

But some older AI studies were still citation magnets in 2023. One paper³ published in 1997 – which describes an early neural-network architecture for modelling

languages called long short-term memory (LSTM) – was the tenth most cited in 2023. Because of their efficiency, LSTMs remain popular for processing some data. Jürgen Schmidhuber, a pioneer of AI and one of the authors of the 1997 paper, who is now at King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, says that not all old AI papers get the same recognition, in part because those articles put forward ideas before they were feasible.


The citations just grew and grew and grew."

"Sadly, the field of AI and machine learning is rife with plagiarism," he says, "and some of the most famous papers actually failed to cite the original work published when compute was millions of times more expensive than today."

Materials high

Remarkably, another research paper⁴ written nearly three decades ago was the fourth most referenced work in papers published in 2023. In 1996, three researchers at Tulane University in New Orleans, Louisiana, published a clever, fast approximation that could be used in software to help researchers to calculate the interactions of electrons in materials, as a way to understand the materials' properties.

"The citations just grew and grew and grew," says physicist Kieron Burke, who co-authored the article with physicists John Perdew and Matthias Ernzerhof. In fact, one-quarter of the paper's total citations were garnered in the

past two years, according to the Dimensions research database, and it is also the fourth most-cited paper of all time (see page 591).

In the past decade, says Burke, who is now at the University of California, Irvine, universities and companies all over the world have been investing in ways to explore new advanced materials – and that's led to a surge in use of the algorithms involved. "These electronic-structure calculations are key to the future of materials design," he says.

The study is one of many in the field of density functional theory (DFT), a way of simplifying the equations of quantum mechanics when applied to molecules and materials. Another DFT paper⁵, also published in 1996, is the ninth most cited in 2023 and eighth of all time.

Burke doubts many researchers are reading the original papers – but thinks that a culture has built up of citing key DFT source methods. That's in part, he thinks, because there are several competing approximation algorithms in the field, so it is unusually important to be specific about the method used.

Measurement analysis

The oldest paper⁶ still cited heavily today – in the top 20 – is an article published in 1981. It discusses a statistical method to tackle issues with modelling relationships between variables that researchers measure, and their 'latent' factors, which are suspected but not observed. Measurements of people's results on cognitive tests might be related to underlying 'intelligence' or 'personality', for instance.

The paper came out at a time when researchers were using latent-variable analysis in fields such as psychology, marketing and sociology, says co-author David Larcker, who studies accounting and corporate governance at Stanford University in California. "As the use of latent-variable analysis exploded over time, our measure became something that many researchers regularly reported in their papers."

Richard Van Noorden is a features editor with *Nature* in London.

Additional reporting by Matthew Hutson.

1. Thor, A., Marx, W., Leydesdorff, L. & Bornmann, L. Preprint at arXiv <https://doi.org/10.48550/arXiv.1601.01199> (2016).
2. Haunschild, R., Thor, A. & Bornmann, L. Preprint at Edmond <https://doi.org/10.17617/3.B41MPK> (2025).
3. Hochreiter, S. & Schmidhuber, J. *Neural Comput.* **9**, 1735–1780 (1997).
4. Perdew, J. P., Burke, K. & Ernzerhof, M. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
5. Kresse, G. & Furthmüller, J. *Phys. Rev. B* **54**, 11169 (1996).
6. Fornell, C. & Larcker, D. F. *J. Mark. Res.* **18**, 39–50 (1981).

Supplementary information accompanies this article online (see go.nature.com/4j5xtfh).