

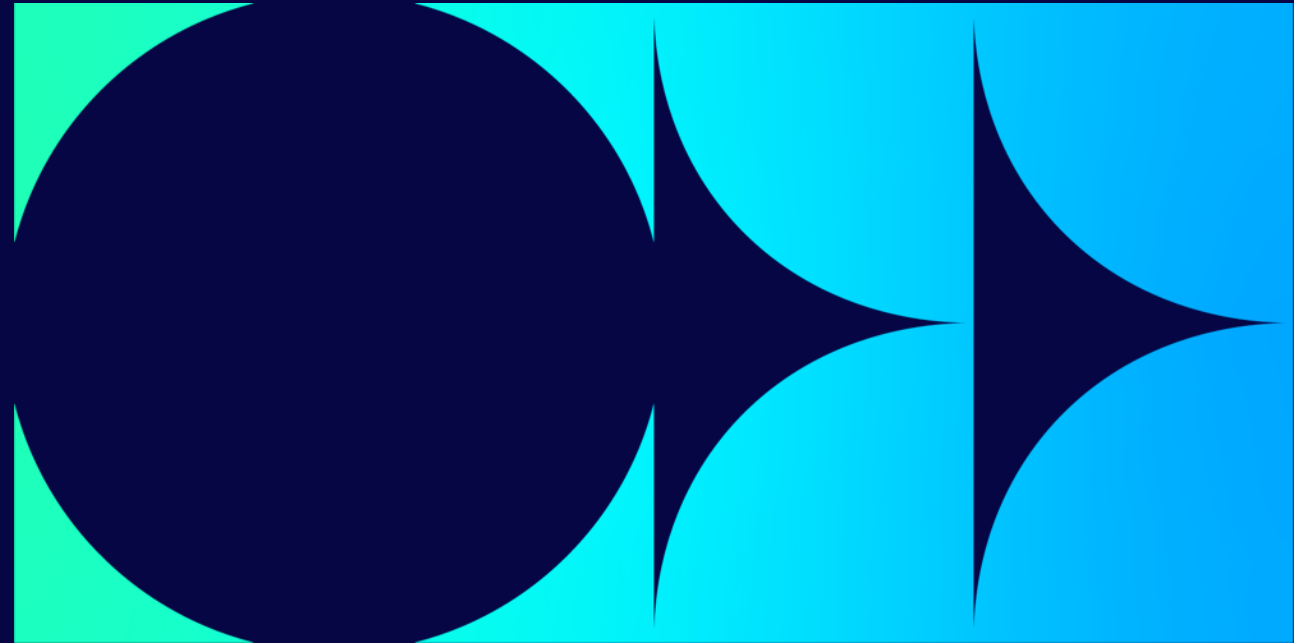
# Agent Ref

An automatic reference checker to support ALs  
(and hence support students)

**Mark Hintze, Janette Wallace and Karen New**

LHCS

30<sup>th</sup> April 2026



# Setting the scene

## S390 capstone module

- 3 TMAs (400 words, 800 words, 1800 words) and EMA (5000- word investigative report/Lit review)
- All projects include critical review of literature
- Tutors to award marks for referencing and citations and support students with skill
- Highlighting any issues to MT where AL see issues

### **Problem:**

Supporting students with referencing skills. And MT ask ALs to check references. Especially important in TMA03/EMA

- Time consuming for ALs to check all reference list
- So ALs make random checks
- Could be missing incorrect/fabricated references.
- Unable to fully support students in developing referencing skills

# Referencing challenges

## Students may not fully understand *why* they are referencing

- References seen as *ornamental* rather than integral to argument
- Believing referencing is about "*avoiding plagiarism*" not engaging scholarship
- Not understanding referencing to position themselves within the academic conversation

## *Cognitive overload* while learning academic writing

- Understanding new content
- Structuring an argument
- Referencing becomes a *secondary task*, done quickly at the end when mental resources are depleted

## **We all want to support students and help them to develop and understanding of appropriate referencing (the 'ingredients' as well as the 'format')**

- Scaffold learning
- Identify errors – and model good practice

# The problem: checking student references

Berry, A.S.F. et al. (2021) 'An Open-Source Toolkit To Expand Bioinformatics Training in Infectious Diseases.', *mBio*, 12(4), p. e0121421. [doi:/10.1128/mBio.01214-21](https://doi.org/10.1128/mBio.01214-21).

Kostopoulou, O., Rosen, A., Round, T., Wright, E., Douiri, A. and Delaney, B. (2017) 'Diagnostic accuracy of GPs when using an early-intervention decision support system: a high-fidelity simulation', *British Journal of General Practice*, 67(656), pp. e201-e208. doi: 10.3399/bjgp17X689233.

Çe, N. and Kumpf, T. (2024) 'Understanding Functional Neurological Disorder: Recent Insights and Diagnostic Challenges', *Journal of Clinical Medicine*, 13(10), p. 2970. doi: 10.3390/jcm13102970.

Which one is real?

Which parts are wrong?

## References

Kostopoulou, O., Rosen, A., Round, T., Wright, E., Douiri, A. and Delaney, B. (2017) 'Diagnostic accuracy of GPs when using an early-intervention decision support system: a high-fidelity simulation', *British Journal of General Practice*, 67(656), pp. e201-e208. doi: 10.3399/bjgp17X689233.

Goyder, C. et al. (2015) 'Diagnostic decision-making and strategies to improve diagnosis in general practice: a qualitative study with newly qualified and experienced doctors', *British Journal of General Practice*, 65(641), pp. e796-e803. doi: 10.3399/bjgp15X687601.

Berry, A.S.F. et al. (2021) 'An Open-Source Toolkit To Expand Bioinformatics Training in Infectious Diseases.', *mBio*, 12(4), p. e0121421. Available at: <https://doi.org/10.1128/mBio.01214-21>.

Borchers, C.H. et al. (2014) 'The Human Proteome Organization Chromosome 6 Consortium: integrating chromosome-centric and biology/disease driven strategies.', *Journal of proteomics*, 100, pp. 60–67. Available at: <https://doi.org/10.1016/j.jpro.2013.08.001>.

Bell, J., Harris, K., Potts, L. and Young, R., 2019. Socioeconomic factors and opioid-related deaths in the UK. *Journal of Public Health*, 41(3), pp.456–462. <https://doi.org/10.1093/pubmed/ffy1234>

Calderone, A. et al. (2016) 'Comparing Alzheimer's and Parkinson's diseases networks using graph communities structure.', *BMC systems biology*, 10, p. 25. Available at: <https://doi.org/10.1186/s12918-016-0270-7>.

Chandran, M. et al. (2023) 'Defining atherosclerotic plaque biology by mass spectrometry-based omics approaches.', *Molecular omics*, 19(1), pp. 6–26. Available at: <https://doi.org/10.1039/d2mo00260d>.

Choi, H., Song, W.-M. and Zhang, B. (2017) 'Linking childhood allergic asthma phenotypes with endotype through integrated systems biology: current evidence and research needs.', *Reviews on environmental health*, 32(1–2), pp. 55–63. Available at: <https://doi.org/10.1515/reveh-2016-0054>.

Choi, S. et al. (2014) 'Procedures for the biochemical enrichment and proteomic analysis of the cytoskeleton.', *Analytical biochemistry*, 446, pp. 102–107. Available at: <https://doi.org/10.1016/j.ab.2013.10.025>.

Cortese, N. et al. (2024) 'Applications of genome-scale metabolic models to the study of human diseases: A systematic review.', *Computer methods and programs in biomedicine*, 256, p. 108397. Available at: <https://doi.org/10.1016/j.cmpb.2024.108397>.

Cuestas, E. (2007) '[Origins of health and disease].', *Revista de la Facultad de Ciencias Médicas (Cordoba, Argentina)*, 64(3), pp. 73–86.

De Gregorio, E. and Rappuoli, R. (2012) 'Vaccines for the future: learning from human immunology.', *Microbial biotechnology*, 5(2), pp. 149–155. Available at: <https://doi.org/10.1111/j.1751-7915.2011.00276.x>.

Deng, L. et al. (2019) 'Fusion of multiple heterogeneous networks for predicting circRNA-disease associations.', *Scientific reports*, 9(1), p. 9605. Available at: <https://doi.org/10.1038/s41598-019-45954-x>.

Gong, L. et al. (2013) 'Prioritization of Disease Susceptibility Genes Using LSM/SVD', *IEEE transactions on bio-medical engineering*, 60(12), pp. 3410–3417. Available at: <https://doi.org/10.1109/TBME.2013.2257767>.

He, F. et al. (2017) 'PCID: A Novel Approach for Predicting Disease Comorbidity by Integrating Multi-Scale Data', *IEEE/ACM transactions on computational biology and bioinformatics*, 14(3), pp. 678–686. Available at: <https://doi.org/10.1109/TCBB.2016.2550443>.

Hicks, J. (2003) 'Genome, proteome, and metabolome: where are we going?', *Ultrastructural pathology*, 27(5), pp. 289–294.

Himmelfarb, J. and Shankland, S.J. (2009) 'Creating research infrastructure and functionality to address chronic kidney disease: the Kidney Research Institute.', *Seminars in nephrology*, 29(5), pp. 457–466. Available at: <https://doi.org/10.1016/j.semnephrol.2009.06.002>.

Jalili, M. et al. (2016) 'Cancerome: A hidden informative subnetwork of the diseaseome.', *Computers in biology and medicine*, 76, pp. 173–177. Available at: <https://doi.org/10.1016/j.complbiomed.2016.07.010>.

Kenah, E. et al. (2016) 'Molecular Infectious Disease Epidemiology: Survival Analysis and Algorithms Linking Phylogenies to Transmission Trees.', *PLoS computational biology*, 12(4), p. e1004869. Available at: <https://doi.org/10.1371/journal.pcbi.1004869>.

Kreek, M.J. et al. (2004) 'Evolving perspectives on neurobiological research on the addictions: celebration of the 30th anniversary of NIDA.', *Neuropharmacology*, 47 Suppl 1, pp. 324–344. Available at: <https://doi.org/10.1016/j.neuropharm.2004.07.024>.

Schaefer, R. et al. (2013) 'Specific collaborative group intervention for patients with medically unexplained symptoms in general practice: a cluster randomized controlled trial', *Psychotherapy and Psychosomatics*, 82(2), pp. 106–119. doi: 10.1159/000342713.

Kuan, X.-Y. et al. (2023) 'Exploring the Causal Relationship Between Telomere Biology and Alzheimer's Disease.', *Molecular neurobiology*, 60(8), pp. 4169–4183. Available at: <https://doi.org/10.1007/s12035-023-03337-4>.

Kumari, R. et al. (2022) 'mRNA modifications in cardiovascular biology and disease: with a focus on m6A modification.', *Cardiovascular research*, 118(7), pp. 1680–1692. Available at: <https://doi.org/10.1093/cvr/cvab160>.

Lewis, E. et al. (2015) 'Carp edema virus/Koi sleepy disease: an emerging disease in Central-East Europe.', *Transboundary and emerging diseases*, 62(1), pp. 6–12. Available at: <https://doi.org/10.1111/tbed.12293>.

Li, H. and Auwerx, J. (2020) 'Mouse Systems Genetics as a Prelude to Precision Medicine.', *Trends in genetics*: TIG, 36(4), pp. 259–272. Available at: <https://doi.org/10.1016/j.tig.2020.01.004>.

Liu, Y. et al. (2025) 'TargetRX Atlas: a database of experimentally supported mRNAs and drug targets in disease.', *Computers in biology and medicine*, 198(Pt A), p. 111191. Available at: <https://doi.org/10.1016/j.complbiomed.2025.111191>.

Lorenzo-Pouso, A.I. et al. (2018) 'Protein-Based Salivary Profiles as Novel Biomarkers for Oral Diseases.', *Disease markers*, 2018, p. 6141845. Available at: <https://doi.org/10.1155/2018/6141845>.

Ma, Y. et al. (2022) 'LncRNA: a new perspective on the study of neurological diseases.', *Biochemical Society transactions*, 50(2), pp. 951–963. Available at: <https://doi.org/10.1042/BST20211181>.

Martínez-García, M. and Hernández-Lemus, E. (2024) 'The Molecular Comorbidity Network of Periodontal Disease.', *International journal of molecular sciences*, 25(18). Available at: <https://doi.org/10.3390/ijms251810161>.

Mehdizadeh, M. et al. (2022) 'The role of cellular senescence in cardiac disease: basic biology and clinical relevance.', *Nature reviews. Cardiology*, 19(4), pp. 250–264. Available at: <https://doi.org/10.1038/s41569-021-00624-2>.

Moore, J.B. (2019) 'From sugar to liver fat and public health: systems biology driven studies in understanding non-alcoholic fatty liver disease pathogenesis.', *The Proceedings of the Nutrition Society*, 78(3), pp. 290–304. Available at: <https://doi.org/10.1017/S0029665119000570>.

Näf, D. et al. (2002) 'The Mouse Tumor Biology Database: a public resource for cancer genetics and pathology of the mouse.', *Cancer research*, 62(5), pp. 1235–1240.

Nichele, S. et al. (2023) 'Hematopoietic cell transplantation for telomere biology diseases: A retrospective single-center cohort study.', *European journal of haematology*, 111(3), pp. 423–431. Available at: <https://doi.org/10.1111/ejh.14023>.

Paplomatas, P., Nikolidaki, M. and Vrahatis, A. (2026) 'Uncovering Key Genes in Neurodegenerative Diseases Through Unsupervised Learning: A Variance-Based LSTM and Enrichment Analysis Approach.', *Advances in experimental medicine and biology*, 1487, pp. 559–568. Available at: [https://doi.org/10.1007/978-3-032-03398-7\\_51](https://doi.org/10.1007/978-3-032-03398-7_51).

Posey, J.E. et al. (2019) 'Insights into genetics, human biology and disease gleaned from family based genomic studies.', *Genetics in medicine: official journal of the American College of Medical Genetics*, 21(4), pp. 798–812. Available at: <https://doi.org/10.1038/s41436-018-0408-7>.

Rende, D., Baysal, N. and Kırdar, B. (2011) 'A novel integrative network approach to understand the interplay between cardiovascular disease and other complex disorders.', *Molecular bioSystems*, 7(7), pp. 2205–2219. Available at: <https://doi.org/10.1039/c1mb05064h>.

Savarino, S.J. and Bourgeois, A.L. (1993) 'Diarrhoeal disease: current concepts and future challenges. Epidemiology of diarrhoeal diseases in developed countries.', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 87 Suppl 3, pp. 7–11. Available at: [https://doi.org/10.1016/0035-9203\(93\)90529-y](https://doi.org/10.1016/0035-9203(93)90529-y).

# How does Agent Ref Work

- ▶ Breaks down a reference into parts: Authors, title, year, journal identifiers and a DOI
- ▶ Searches for the references on Crossref, doi and Pubmed, then matches parts to check the validity of the reference









# Agent Ref Scoring

- Checks whether key citation details agree with a real record, such as authors, title, journal, year and DOI.
- Gives more weight to strong identifiers like a valid DOI but does not rely on the DOI alone.
- Uses a cautious approach: a high score supports confidence, but major mismatches can still flag a reference as suspicious.

## How scoring works

- Combines multiple signals, not just one field.
- Prioritises strong matches such as DOI, title and author agreement.
- Flags references when the details do not fit together, even if part of the citation looks real.

# Agent Ref results

-  Real → DOI verified with resolver + bibliographic match; OR Crossref-derived DOI verified at doi.org + match; OR PMCID/PMID verification
-  Manual review – web source / report / dataset → organisational or dataset/report-style web source with URL
-  Web Source (trusted) /  Web Source → general website/URL
-  Suspicious → partial mismatch or unverifiable but well-formed
-  possible falsification → fabricated pairing or failed checks

# Try it yourself

## Scan the QR code or click the URL

- ▶ Open the test page
- ▶ <https://melted-lily-2ac.notion.site/Agent-Ref-workshop-test-set-f6ad728579114ff69c51060b2a2f6ba6>
- ▶ Copy references
- ▶ Paste into Agent Ref

Which references could you spot by eye and where did the tool help?

Where can you see a use for this tool?



# Potential Future development

## AI Support

- › Open the tool up to use on all modules for more AIs
- › Time saving for AIs and cost savings for the University

## Teaching

- › Create a student teaching tool for use at all levels, to explain referencing issues to students in an interactive.
- › Speed up marking of references

## Research

- › A reference curator, a way of gathering links from academic papers quickly.
- › Save academic time

## Future versions

- › Development of an enhanced version (002) that can analyse student documents and data files

# Summary

## Agent Ref

- › An automatic reference tool
- › Identifies references which don't meet the cite them right format, quickly
- › A tool to speed up a slow and laborious process and provide a rigorous output
- › Potential uses in teaching, marking, academic integrity and research

Thanks for taking part and providing feedback

Please contact: [Janette.Wallace@open.ac.uk](mailto:Janette.Wallace@open.ac.uk), [Karen.new@open.ac.uk](mailto:Karen.new@open.ac.uk),  
and/or [Mark.Hintze@open.ac.uk](mailto:Mark.Hintze@open.ac.uk) for further information

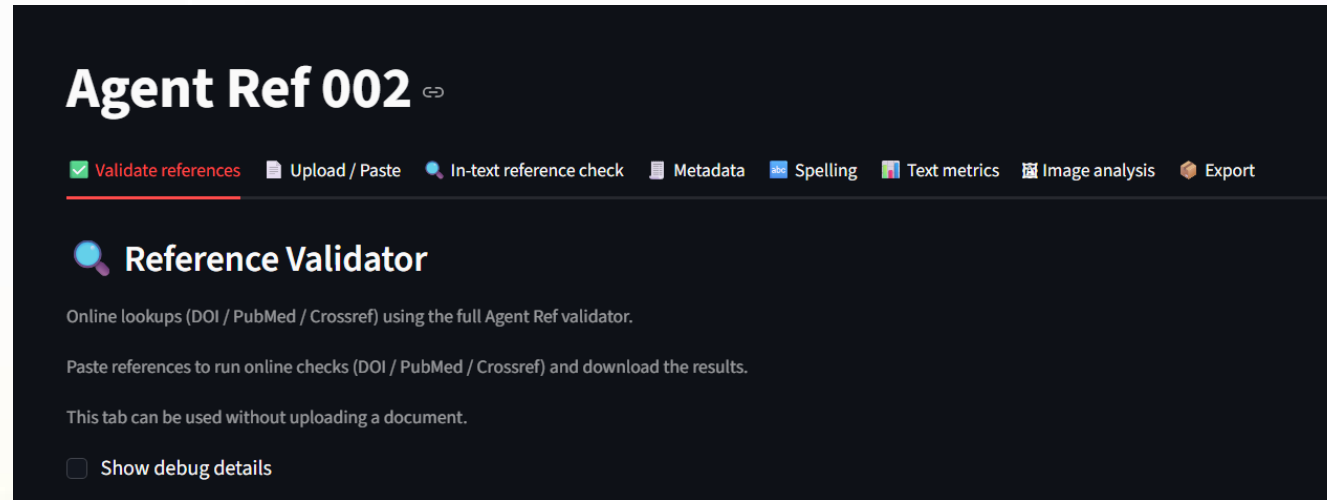


**The Open  
University**

# Thank you

# Agent Ref 002

An extended version intended to work with documents from students



- **Implementations for the future**
- Citation number
- Open University databases – for OU refs and others such as arXiv
- Journal validity
- Link source checking: AI Tags
- Highlight double references
- Academic integrity
- Can check for many red flags in a students work automatically
- Streamline procedure