Zootaxa 5270 (2): 347–350
https://www.mapress.com/zt/

Correspondence

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)
https://doi.org/10.11646/zootaxa.5270.2.12
http://zoobank.org/urn:lsid:zoobank.org:pub:64CD2C27-12EF-4896-9974-B1F361281BAB

Can ChatGPT be leveraged for taxonomic investigations? Potential and limitations of a new technology

ANDREW A. DAVINACK
Biology Department, Wheaton College, 26 East Main Street, Norton, Massachusetts, USA 02766
davinack_drew@wheatoncollege.edu; *https://orcid.org/0000-0002-1263-7119*

Biodiversity assessment surveys are crucial for understanding the impacts of anthropogenic-mediated change on life on our planet (Pauchard *et al.* 2018). Such surveys have been used for decades by researchers to build checklists and inventories of the planet's biological heritage, map the distribution of rare and endemic species and manage biological invasions (Pysek *et al.* 2013; Hochkirch *et al.* 2021). The accuracy of these assessments is largely dependent on the work of dedicated taxonomists whose knowledge have aided in the description and diagnoses of millions of plant and animal species for more than 250 years (Khuroo *et al.* 2007). Nowhere is the need for taxonomists more pressing than in the marine realm which harbors a significant share of life's morphological and genetic diversity (Armbrust & Palumbi, 2015). Yet, despite this need, there is still a severe shortage of taxonomists for many groups with the discovery rate of undescribed and problematic taxa outpacing our ability to train competent taxonomists (Drew 2011). An enterprise that is in urgent need for taxonomists is the rapid assessment survey, which is used to detect non-indigenous species in a timely manner. These early detections provide environmental authorities a large-enough window to implement management and or eradication protocols if needed. Rapid assessment surveys occur across a relatively wide geographic region in a short amount of time, usually one hour of sampling per site. Organisms collected are then sorted and identified to the species level and then inventoried. Over multiple years, such surveys can provide valuable data into the number of alien fauna being transported to a particular region over time (Rohde *et al.* 2017; Pederson *et al.* 2021). Recruiting a complete team of taxonomists for these surveys have historically been difficult simply due to the lack of experts available for specific groups and also due to logistical and financial challenges, especially if there are only a handful of experts and they are all located a considerable distance from the study area. Instead, non-taxonomists often bravely take on the challenge of sorting through various keys (some of which may be outdated or created for a completely different geographic region) and redescriptions (many of which involve specialized terminology only used among specialists). Worse yet, general field guides are sometimes employed resulting in inaccurate identifications, especially for cryptic species complexes, of which there are many within the marine invertebrate realm. These issues ultimately result in the production of incomplete species lists where some individuals can only be identified to the genus or family level. While DNA barcoding can be integrated into the survey, this process takes time, which nullifies the 'rapid' benefits of the survey. Ultimately, many specimens remain unidentified and eventually end up in private collections or community vouchers.

With the rise of large language models (LLMs) that can autonomously learn by being trained on a vast amount of information in real time, there are now growing conversations on whether these artificial intelligence (AI) models can be leveraged for scientific research, and if so, whether they should (van Dis *et al.* 2023; Arif *et al.* 2023; Zhu *et al.* 2023). The most exciting of these LLMs today is undoubtedly the chatbot known as ChatGPT, which was developed by the company OpenAI, based in San Francisco, USA. For science as a whole, ChatGPT is a disruptive technology and has the potential to revolutionize the way that we conceive of studies, run experiments and analyze data. Currently, no formal ethical guidelines exist for using LLMs in scientific research and in many ways, we are in the midst of the 'wild west' of this technology. Thus far, concerns about the technology have been largely directed to its use in scientific writing; summarizing research papers, drafting parts of manuscripts along with its use in statistical analyses and coding. In fact, I have used ChatGPT on numerous occasions to debug python error codes and it has performed almost flawlessly, producing outputs with the correct script which can simply be copied and pasted into the command line of my machine. In an excellent overview of the technology, van Dis *et al.* (2023) cautioned scientists about using LLMs for nuanced tasks that require a high level of cognitive flexibility. In their overview, the authors presented ChatGPT with questions and assignments pertaining to their field of clinical psychology and found that the chatbot produced false and misleading

Accepted by Z.- Q. Zhang: 12 Apr. 2023; published: 20 Apr. 2023

information. As far as I know, there has thus far been no open discussion about whether the technology can be leveraged for taxomomic research.

On the surface, ChatGPT seems as though it could be the solution to the problem of a dearth of taxonomists for biodiversity surveys, specifically during biodiversity surveys (RAS). During these surveys, non-experts ('generalists') who tentatively identifies two specimens that look very similar and clearly belong to the same genus,but is unable to make a definitive confirmation due to some ambiguous character(s) should in theory be able to present ChatGPT with their incomplete diagnoses and have the model produce a final confirmation or at least a clarification of the ambiguity. To test this, I presented a series of queries and follow-up questions to ChatGPT regarding the identity of two shell-boring polychaete worms of the genus *Polydora* Bosc, 1802; notorious pests of commercially reared shellfish such as scallops, oysters and abalone. For the first trial, I wanted to determine whether ChatGPT could accurately distinguish the superficial morphological differences between *Polydora websteri* Hartman in Loosanoff & Engle, 1943 and *Polydora neocaeca* Williams & Radashevsky, 1999—both of which appear identical to a non-polychaetologist looking through a standard compound microscope. For practical purposes, pigmentation patterns and the structure and arrangement of modified spines on the worms' fifth segment (chaetiger) were chosen since they can be used as a diagnostic tool to differentiate the two species. The input and output for the first query is shown below:

Query 1: Provide a detailed overview of the pigmentation differences between *Polydora websteri* and *Polydora neocaeca*.

ChatGPT first provided a general statement of the fact that external morphology is similar between the two species making it difficult to differentiate and that pigmentation could be a diagnostic tool (both of which are correct). The following output was received:

ChatGPT: ["*Polydora websteri* typically has a brownish-gray body with a mottled or marbled appearance caused by irregular patches of black pigment on the dorsal surface. The ventral surface is usually paler, with a white or grayish coloration. The head is also darkly pigmented with two black spots. In contrast, *Polydora neocaeca* has a more uniform pigmentation pattern, with a dark brown or black body that lacks the mottled appearance seen in *P. websteri*. The head is also uniformly darkly pigmented, without the two black spots as seen in *P. websteri.* It is important to note however that pigmentation patterns vary among individuals and may not always be reliable for species identification."]

We were surprised at the accuracy of this description, especially the last line which cautions the user about the problem of pigmentation variability between the two species. We then followed up this with a second query.

Query 2: Describe the fifth chaetiger spines of *P. websteri* and *P. neocaeca* for the purpose of species differentiation.

The abbreviated output is shown below:

ChatGPT: ["The fifth chaetiger of *P. websteri* and *P. neocaeca* can be used to differentiate the two species. *P. websteri* has 3 pairs of spines on the fifth chaetiger…..*P. neocaeca* on the other hand has only 2 pairs of spines on the fifth chaetiger. The differences in the fifth chaetiger spines can be useful in distinguishing between *P. websteri* and *P. neocaeca*, especially where other diagnostic features are not readily apparent. It is important to note, however that accurate identification requires careful examination of multiple characters, including the shape and size of the prostomium, peristomium and parapodia, as well as the number and arrangement of the chaetae"]

Here, ChatGPT produced a critical error in its diagnosis: *P. neocaeca* actually has 7 modified spines on the fifth chaetiger (Malan *et al.* 2020) while *P. websteri* can have 5–7 spines (Martinelli *et al*. 2020). These descriptions are consistent in both the older and the more recent literature on the species. On a positive note, like the previous query, ChatGPT did provide a useful caveat at the end mentioning that complete delineation would require investigating other characters (prostomium, peristomium, chaetal arrangement), all of which are indeed diagnostic for these two species.

For our final query, we asked ChatGPT to produce a list of references to support their outputs. This was where the most egregious problem arose. Of the six references that were provided to support the conclusions, five were fabricated. These fabricated references appeared to be legitimate sources—they were 'authored' by well-respected polychaete taxonomists with familiar-sounding titles and 'published' in well-known taxonomic-focused journals such as *Zootaxa* and *Journal of Natural History*. However, a closer investigation using a combination of Google Scholar, Web of Science, and the journals' archives themselves, failed to find any of the listed references. A study by Zhu *et al.* (2023) looked at the usefulness of ChatGPT to conduct literature reviews in the pharmacy field and they also found that three of the four references provided for a particular query were also fabricated. It appears that these references are the result of the algorithm stitching together various pieces of information regarding prior queries rather than recovering actual studies.

In summary, it appears that ChatGPT and LLMs in general may hold some promise for taxonomic research, however their information distilling capabilities will need to be ironed out to reduce inaccuracies. The limited ability to filter relevant information appears to be an issue with ChatGPT (van Dis *et al.* 2023) which is a serious problem if it is to be applied to the field of taxonomy which not only utilizes the most recent species descriptions but does so while contextualizing older literature. One would hope that this dynamic utilization of data will eventually be mastered by LLMs over time. There are other issues we noticed that for the sake of brevity, were not presented in detail here. For example, when prompted to provide the global distribution of *P. websteri*, less than 50% of its actual distribution showed up in the results. We also presented ChatGPT with a simple query: "Provide a morphological diagnosis of the polychaete *Polydora hoplura*", and while the model did provide accurate information, the characters chosen were not necessary diagnostic (e.g., coloration, size, number of chaetigers). In other words, ChatGPT in its current form cannot be a reliable substitute for expert taxonomists if the goal is to carry out a complete, comprehensive and accurate biodiversity assessment. Instead, the model in its current iteration could be used by non-taxonomists to improve their initial identifications but a review by an actual taxonomist is absolutely necessary before any final confirmations can be made. Due to the aforementioned limitations, this obviously means that the model cannot be used for species discovery, yet. In their review of the technology and its promise for scientific studies, van Dis *et al.* (2023) rightly advocated that scientific funding organizations, NGOs and universities should prioritize the funding of more open-AI learning models where the training sets are publicly available. This will not only make the LLMs more transparent but can also open the door to field-specific LLMs such as a taxonomy/biodiversity LLM where it is actually designed to produce accurate descriptions and detailed diagnoses. Fortunately, research into the use of machine learning for taxonomic research is already underway (Allken *et al.* 2018, Tan *et al.* 2020, Tan *et al.* 2021, Thenmozhi *et al.* 2021), which can provide a model for an open-AI system. Furthermore, investment into image-based learning models (IBLMs) that can be coupled to such LLMs would be the ideal setup to advance this technology to the service of taxonomy. I can envision that this would be of particular interest to national museums where much of their inventory remains unidentified and undescribed and where staffing and curatorial staff shortages abound. In addition, organizations such as the International Barcode of Life would also benefit from a taxonomy-basedAI learning model which automates the descriptive phase of the barcoding workflow, thereby accelerating the cataloguing process.

The molecular revolution transformed the field of taxonomy not too long ago by providing an additional tool for describing and cataloging Earth's biodiversity. The implementation of artificial intelligence may very well be the next frontier for this science should all the right stars align: funding, buy-in from taxonomists and natural history museums and improvements in learning algorithms.

References

Allken V., Handegard, N.O., Rosen, S., Schreyeck, T., Mahiout, T. & Malde, K. (2018) Fish species identification using a convolutional neural network trained on synthetic data. *ICES Journal of Marine Science,* 76 (1), 342–349. https://doi.org/10.1093/icesjms/fsy147

Arif T.B., Munaf, U. & Ul-Haque, I. (2023) The future of medical education and research: Is ChatGPT a blessing or blight in disguise? *Medical education online,* 28 (1), 2181052. https://doi.org/10.1080/10872981.2023.2181052

Armbrust, E.V. & Palumbi, S.R. (2015) Uncovering hidden worlds of ocean biodiversity. *Science*, 348 (6237), 865–867. https://doi.org/10.1126/science.aaa7378

Drew, L.W. (2011) Are we losing the science of taxonomy? As need grows, numbers and training are failing to keep up. *BioScience*, 61 (12), 942–946. https://doi.org/10.1525/bio.2011.61.12.4

Hochkirch, A., Samways, M.J., Gerlach, J., Bohm, M., Williams, P., Cardoso, P., Cumberlidge, N., Stephenson, P.J., Seddon, M.B., Clausnitzer, V., Borges, P.A.V., Mueller, G.M., Pearce-Kelly, P., Raimondo, D.C., Danielczak, A. & Dijkstra, K-D.B. (2021) A strategy for the next decade to address data deficiency in neglected biodiversity. *Conservation Biology*, 35 (2), 502–509. https://doi.org/10.1111/cobi.13589

Khuroo, A.A., Dar, G.H., Khan, Z.S. & Malik, A.H. (2007) Exploring an inherent interface between taxonomy and biodiversity: current problems and future challenges. *Journal for Nature Conservation*, 15, 256–261. https://doi.org/10.1016/j.jnc.2007.07.003

Malan, A., Williams, J.D., Abe, H., Sato-Okoshi, W., Matthee, C.A. & Simon, C.A. (2020) Clarifying the cryptogenic species *Polydora neocaeca* Williams & Radashevsky, 1999 (Annelida: Spionidae): a shell-boring invasive pest of molluscs from locations worldwide. *Marine Biodiversity*, 50, 51. https://doi.org/10.1007/s12526-020-01066-8

Martinelli, J.C., Lopes, H.M., Hauser, L., Jimenez-Hidalgo, I., King, T.L., Padilla-Gamino, J.L., Rawson, P., Spencer, L.H., Williams, J.D. & Wood, C.L. (2020) Confirmation of the shell-boring oyster parasite *Polydora websteri* (Polychaeta: Spionidae) in Washington State, USA. *Scientific Reports*, 10, 3961.
https://doi.org/10.1038/s41598-020-60805-w

Pauchard, A., Meyerson, L.A., Bacher, S., Blackburn, T.M., Brundu, G., Cadotte, M.W., Courchamp, F., Essl, F., Genovesi, P., Haider, S., Holmes, N.D., Hulme, P.E., Jeschke, J.M., Lockwood, J.L., Novoa, A., Nunez, M.A., Peltzer, D.A., Pysek, P., Richardson, D.M., Simberloff, D., Smith, K., van Wilgen, B.W., Vila, M., Wilson, J.R.U., Winter, M. & Zenni, R.D. (2018) Biodiversity assessments: origin matters. *PLoS Biology*, 16 (11), e2006686.
https://doi.org/10.1371/journal.pbio.2006686

Pederson, J., Carlton, J.T., Bastidas, C., David, A., Grady, S., Green-Gavrielidis, L., Hobbs, N-V., Kennedy, C., Knack, J., McCuller, M., O'Brien, B., Osborne, K., Pankey, S. & Trott, T. (2021) *BioInvasions Records*, 10 (2), 227 – 237.
https://doi.org/10.3391/bir.2021.10.2.01

Pysek, P., Hulme, P.E., Meyerson, L.A., Smith, G.F., Boatwright, J.S., Crouch, N.R., Figueiredo, E., Foxcroft, L.C., Jarosik, V., Richardson, D.M., Suda, J. & Wilson, J.R.U. (2013) Hitting the right target: taxonomic challenges for, and of, plant invasions. *AoB PLANTS,* 5.
https://doi.org/10.1093/aobpla/plt042

Rohde, S., Schupp, P.J., Markert, A. & Wehrmann, A. (2017) Only half of the truth: Managing invasive alien species by rapid assessment. *Ocean & Coastal Management*, 146, 26–35.
https://doi.org/10.1016/j.ocecoaman.2017.05.013

Thenmozhi, K., Dakshayani, S. & Srinivasulu, R.U. (2021) Insect classification and detection in field crops using modern machine learning techniques. *Information Processing in Agriculture,* 8, 446–457.
https://doi.org/10.1016/j.inpa.2020.09.006

Tan, J.W., Chang, S.W., Abdul-Kareem, S., Yap, H.J. & Yong, K.T. (2020) Deep learning for plant species classification using leaf vein morphometric. *IEE-ACM Transactions on Computational Biology and Bioinformatics,* 17 (1), 82–90.
https://doi.org/10.1109/TCBB.2018.2848653

Tan, H.Y., Goh, Z.Y., Loh, K., Then, A.Y., Omar, H. & Chang, S. (2021) Cephalopod species identification using integrated analysis of machine learning and deep learning approaches. *PeerJ*, 9, e11825.
https://doi.org/10.7717/peerj.11825

Thomson, S.A., Pyle, R.L., Ahyong, S.T., Alonso-Zarazaga, M., Ammirati, J., Araya, J.F., Ascher J.S., Audisio, T.L., Azevedo-Santos, V.M., Bailly, N., Baker, W.J., Balke, M., Barclay, M.V.L., Barrett, R.L., Benine, R.C., Bickerstaff, J.R.M., Bouchard, P., Bour, R., Bourgoin, T., Boyko, C.B., Breure, A.S.H., Brothers, D.J., Byng, J.W., Campbell, D., Ceríaco, L.M.P., Cernák, I., Cerretti, P., Chang, C.-H., Cho, S., Copus, J.M., Costello, M.J., Cseh, A., Csuzdi, C., Culham, A., D'Elía, G., d'Udekem d'Acoz, C., Daneliya, M.E., Dekker, R., Dickinson, E.C., Dickinson, T.A., van Dijk, P.P., Dijkstra, K.-D.B., Dima, B., Dmitriev, D.A., Duistermaat, L., Dumbacher, J.P., Eiserhardt, W.L., Ekrem, T., Evenhuis, N.L., Faille, A., Fernández-Triana, J.L., Fiesler, E., Fishbein, M., Fordham, B.G., Freitas, A.V.L., Friol, N.R., Fritz, U., Frøslev, T., Funk, V.A., Gaimari, S.D., Garbino, G.S.T., Garraffoni, A.R.S., Geml, J., Gill, A.C., Gray, A., Grazziotin, F.G., Greenslade, P., Gutiérrez, E.E., Harvey, M.S., Hazevoet, C.J., He, K., He, X., Helfer, S., Helgen, K.M., van Heteren, A.H., Hita Garcia, F., Holstein, N., Horváth, M.K., Hovenkamp, P.H., Hwang W.S., Hyvönen, J., Islam, M.B., Iverson, J.B., Ivie, M.A., Jaafar Z., Jackson, M.D., Jayat, J.P., Johnson, N.F., Kaiser, H., Klitgård, B.B., Knapp, D.G., Kojima, J.-I., Kõljalg, U., Kontschán, J., Krell, F.-T., Krisai-Greilhuber, I., Kullander, S., Latella, L., Lattke, J.E., Lencioni, V., Lewis, G.P., Lhano, M.G., Lujan, N.K., Luksenburg, J.A., Mariaux, J., Marinho-Filho, J., Marshall, C.J., Mate, J.F., McDonough, M.M., Michel, E., Miranda, V.F.O., Mitroiu, M.-D., Molinari, J., Monks, S., Moore, A.J., Moratelli, R., Murányi, D., Nakano, T., Nikolaeva, S., Noyes, J., Ohl, M., Oleas, N.H., Orrell, T., Páll-Gergely, B., Pape, T., Papp, V., Parenti, L.R., Patterson, D., Pavlinov, I.Y., Pine, R.H., Poczai, P., Prado, J., Prathapan, D., Rabeler, R.K., Randall, J.E., Rheindt F.E., Rhodin, A.G.J., Rodríguez, S.M., Rogers, D.C., Roque, F.D.O., Rowe, K.C., Ruedas, L.A., Salazar-Bravo, J., Salvador, R.B., Sangster, G., Sarmiento, C.E., Schigel, D.S., Schmidt, S., Schueler, F.W., Segers, H., Snow, N., Souza-Dias, P.G.B., Stals, R., Stenroos, S., Stone, R.D., Sturm, C.F., Štys, P., Teta, P., Thomas, D.C., Timm, R.M., Tindall, B.J., Todd, J.A., Triebel, D., Valdecasas, A.G., Vizzini, A., Vorontsova, M.S., de Vos, J.M., Wagner, P., Watling, L., Weakley, A., Welter-Schultes, F., Whitmore, D., Wilding, N., Will, K., Williams, J., Wilson, K., Winston, J.E., Wüster, W., Yanega, D., Yeates, D.K., Zaher, H., Zhang, G., Zhang, Z.-Q. & Zhou, H.-Z. (2018) Taxonomy based on science is necessary for global conservation. *PLoS Biology*, 16, e2005075.
https://doi.org/10.1371/journal.pbio.2005075

van Dis, E.A.M., Bollen, J., Zuidema, W., van Rooij, R. & Bockting, C.L. (2023) ChatGPT: five priorities for research. *Nature*, 614, 224–226.
https://doi.org/10.1038/d41586-023-00288-7

Zhu, J.J., Jiang, J., Yang, M. & Ren, Z.J. (2023) ChatGPT and environmental research. *Environmental Science & Technology.*
https://doi.org/10.1021/acs.est.3c01818

Zhu, Y., Han, D., Chen, S., Zeng, F. & Wang, C. (2023) How can ChatGPT benefit pharmacy: A case report on review writing. Preprints, Mar. 20, 2023.
https://doi.org/10.20944/preprints202302.0324.v1