

Integrating Vector Control: A Large Language Model Chatbot for Interdisciplinary Design Processes

Otis Sloan Wood¹[0000-0002-5494-9540], Muhammad Farooq²

¹Royal Danish Academy, Copenhagen, Denmark

²Prompt Engineering, Seattle, USA

ABSTRACT: Working across disciplines is essential to tackle the complex global health challenges faced today. One such challenge is the proliferation of *Aedes* mosquitoes in recent decades, which has led to a significant and growing public health burden. According to the WHO, half of the world's population is at risk of contracting *Aedes* mosquito-borne viral diseases (MBVDs) such as dengue, and this is expected to rise as climate change expands the predicted geographic reach of these mosquitoes as far north as Denmark by 2050. There is currently limited vaccines or preventative therapeutics for these MBVDs, which means mosquito control remains the most viable approach to reduce them. Modifications to the built environment present numerous opportunities to reduce this disease burden, such as designing drainage that prevents stagnant water that can become larval sites or utilising screening to prevent mosquito entry. Significant advancements in vector control strategies have been made in the fields of public health and entomology; however, these have not effectively disseminated into the building profession. This paper explores the development and application of a proof-of-concept chatbot platform that utilises Large Language Models (LLMs) to bridge architecture and public health disciplines, aiding the integration of evidence-based mosquito control strategies into the early stages of building and landscape design. This LLM chatbot draws information from peer-reviewed articles from leading global health journals, identified in a previous literature review. Users can easily access information on vector control approaches for their projects by asking simple questions such as, 'How can I improve my entrance space to prevent mosquitoes?'. By curating the LLM's source information, we aim to reduce AI hallucinations, where fictitious generated information is presented as fact. This paper reflects upon the opportunities and challenges of LLM tools for interdisciplinary projects identified during the chatbots development.

KEYWORDS: mosquito-borne diseases, vector control, large language model, global health, building design

1.0 INTRODUCTION

1.1 The increasing global health burden of *Aedes* mosquito-borne viral diseases

Many of the global health challenges faced today, such as infectious diseases and healthcare access, are impacted by factors spanning multiple disciplines, including fields such as epidemiology, public health, environmental science and the built environment. One such health challenge is the increasing risk of mosquito-borne viruses transmitted by *Aedes Aegypti* and *Aedes Albopictus*. Presently around half of the world's population are at risk of contracting one of the multitude of viruses these mosquitoes are vectors of (World Health Organisation, 2023). These diseases include dengue, chikungunya, Zika and yellow fever. Of these diseases, there is only a readily available and universally recommended vaccine for yellow fever.

The population and geographic spread of *Aedes* mosquitoes has risen dramatically in the last two decades. This is due to a variety of factors. Increased global populations, travel, trade, and urban expansion have helped these mosquito species spread across the globe (Shragai et al. 2017). *Aedes* mosquitoes demonstrate significant ecological plasticity and have adapted from laying their eggs in small water-filled tree-holes, to deposit their eggs in a wide range of artificial water bodies found in places of human habitation, such as water storage containers, discarded waste, drains and slumps. Higher temperatures, increased humidity and more erratic precipitation patterns exacerbated by climate change, has led to an increase in *Aedes* mosquito populations, their capacity as disease vectors and habitat range (Watts et al. 2021; Murray et al. 2020). This habitat expansion is expected to increase over the coming decades. A study by Oliveira, S. et al shown in Figure 1, modelling climate suitability for *Aedes albopictus* mosquitoes in 2050, predicts that 47% of Europe will be climatically suitable for this mosquito (Oliveira et al. 2021).

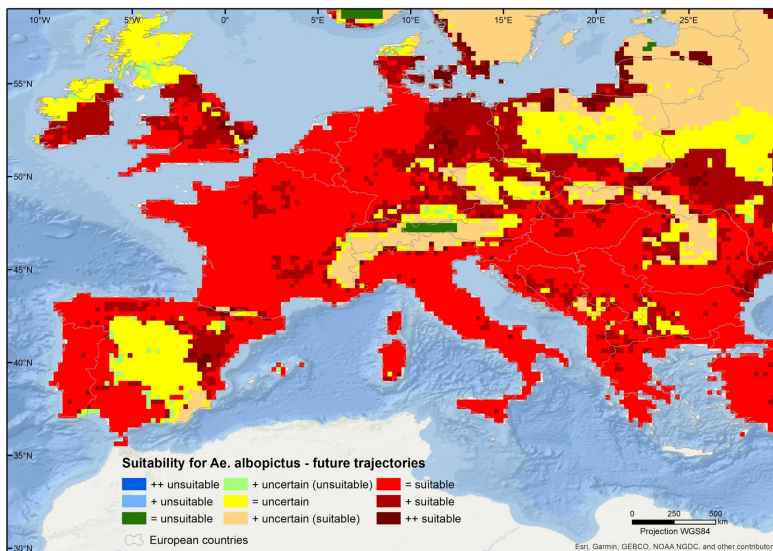


Figure 1: Projected climate suitability of *Aedes albopictus* mosquitoes in 2050. Source: (Oliveira, S. et al. 2021).

1.2 Integrating *Aedes* vector control into building design

Given the growing population and habitat spread of *Aedes* mosquitoes and the limited vaccinations or treatment options for the diseases they transmit, it is crucial to implement effective vector control methods to reduce this escalating public health burden. Preventative vector control measures targeting *Aedes* mosquitoes can be a cost-effective strategy compared to therapeutic interventions or vaccinations, as they prevent the transmission of multiple diseases transmitted by a single vector. Furthermore, many of these measures are also effective against other mosquito vectors, thereby enhancing their overall public health impact. In order to develop these preventative approaches, the WHO has called for integrated and co-creational approach to vector control, bringing together specialists across different disciplines (Achee et al. 2015; WHO 2017). One such approach is the incorporation of vector control methods into the built environment. Recent studies pictured in Figure 2, have utilised building design as means to protect against mosquitoes, including novel screened facades, improved ventilation to reduce the internal concentrations of mosquito attractants such as carbon dioxide, and elevated sleeping spaces (Mshamu et al. 2022; Carrasco-Tenezaca et al. 2021; von Seidlein et al. 2017).



Figure 2: (Left) Exterior view of a Star Home in Tanzania, that utilises a variety of design strategies to reduce the risk of inhabitants contracting Malaria, including screened windows, increased ventilation and raised bedrooms. Source: (Mshamu et al. 2022). (Right) Experimental houses in the Gambia designed to test the relationship between building elevation and mosquito entry. Source: (Carrasco-Tenezaca et al. 2021)

These studies have brought together building professionals, entomologists, global health specialists and communities at risk of MBVs to develop novel approaches for vector control integrated within the built environment. Such approaches can offer other co-benefits such as improved housing conditions. Many regions affected by mosquito-borne viral diseases (MBVDs), such as Sub-Saharan Africa, are expected to experience significant population growth and increased built infrastructure demands over the coming decades. Incorporating vector control measures into future building designs could effectively mitigate the rising threat of *Aedes*-transmitted diseases while meeting the growing needs. Figure 3 highlights the different areas of building design that can be used to reduce the risk of MBVD transmission, identified in a previous literature review by the authors. These include landscape and water management strategies to reduce larval breeding sites, architectural and building component design to block mosquito entry and building user mobilisation to increase awareness and adoption of preventative methods.

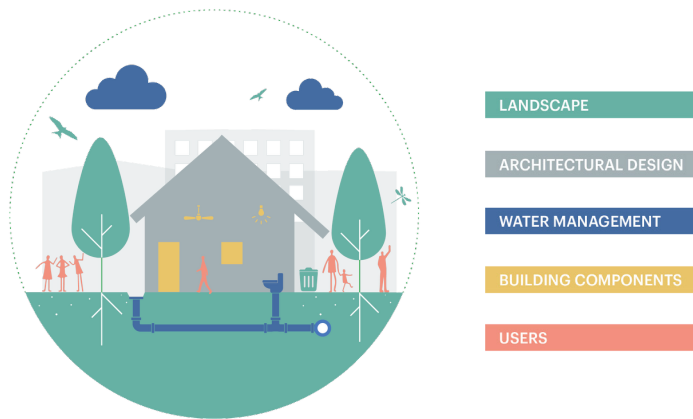


Figure 3: Diagram outlining the different areas within built environment design that can be adapted to reduce the reduce of mosquito-borne diseases. Source: (Author 2023)

1.3 Barriers to accessing vector control information for building professionals

Excluding the earlier studies mentioned, progress in built approaches to vector control has been fairly limited within the architectural discourse. Knowledge and the incorporation of vector control methods, particularly in the early stages of the design process, could dramatically reduce MBVD risk for building users and have the high cost to gain ratio when compared to retrofitting control measures (Obonyo, Pareek, and Woldu 2019). One reason for this lack of uptake may be that academic journals serve as the primary publication medium for vector control interventions, posing access barriers for building designers, contractors, and self-builders due to challenges such as academic writing style, paywalls or lack of integration into the architecture school curriculum. For built environment professionals interested to learn more, a further barrier may be the overabundance of literature available on the topic presents challenges in terms of find relevant information and staying up to date. For example, at the time of writing, a search of articles on '*Aedes aegypti*' using Google Scholar presents over 400,000 results.

1.4 Utilising large language models for information retrieval from text data sources

Large Language Models (LLMs) offer a potential tool to help improve access to relevant information for building professionals on *Aedes* vector control interventions and the latest research on other health burdens related to the built environment. LLMs are a form of artificial intelligence algorithm which use large data sets, typically consisting of text-based articles, combined with deep learning techniques to understand and summarise written information as well as creating new content. LLMs are the primary technology behind ChatGPT, (Generative Pre-trained Transformer) developed by OpenAI that was released in 2018 as well as Google's BERT (Bidirectional Encoder Representations from Transformers). LLMs can work with large text datasets which could include thousands of articles and therefore offer a powerful tool in enabling building professionals to access the wealth of information within public health and entomology literature and translate this into design approaches that could help reduce the risk of *Aedes* transmitted diseases. There are a variety of LLM projects developed by academics to utilise this technology for research purposes, such as BLOOM (BigScience Large Open-science Open-access Multilingual Language Model), an open-source model aimed to provide transparency over its 1.6TB of training data (Scao et al. 2022) or GatorTron, which is trained on over 90 billion words of text from health records, and has been tested to answer health related questions (Yang et al. 2022).

A key barrier for the use of existing AI chatbots such as ChatGPT to provide information on *Aedes* control methods and other research questions, is the likelihood of receiving fictitious results known as 'AI hallucinations'. AI hallucinations are a phenomenon that occurs when an LLM misinterprets patterns within its training data, generating outputs that are not grounded in factual reality. Due to these results being based on existing factual data, these hallucinations can be highly convincing. Another challenge is a lack of control over the source and reference data from which the LLM is trained and draws the information to answer a query. It is often difficult to determine the source of an answer from a chatbot and harder to know whether this has gone through the scrutiny of the peer review process. The lack of interpretability of answers given by LLMs, like the free version of ChatGPT, raise other concerns when interpreting responses, such as the lack of clarity over why certain information was prioritised and present above other information (Kasneji et al. 2023).

This paper describes the development and application of an LLM chatbot platform, designed to help bridge the gaps between architecture and public health domains and enable better integration of evidence-based mosquito control strategies into building and landscape design. This LLM chatbot, nicknamed '*Aedes* Assistant', takes information from a database of select peer-reviewed articles from leading global health journals. Users can then easily access information on vector control strategies appropriate to their projects by asking simple questions such as, '*How can I improve entrance space to prevent mosquitoes?*'. By selecting the LLM's article-based source information, we aim to reduce fictitious AI hallucinations.

2.0 METHOD

2.1 Development of the *Aedes* chatbot

To reduce the likelihood of generated AI hallucinations and ensure responses were based on high quality sources and improve interpretability, we aimed to create a chatbot that was powered by an LLM, such as OpenAI, but used information to answer queries based on a curated knowledge database of peer reviewed journals. An electronic literature review was conducted using the PubMed database with the search term '*Aedes*' and mesh term '*vector control*' or '*insect control*'. From the body of literature available on *Aedes* vector control interventions (3,367 articles), we screened for doubles using Zotero and whether they concerned or could be applied to the built environment. For example, the large body of articles focused on the sterilisation of *aedes* mosquitoes or genetic modification as a means of control were not included. From this, 82 articles were identified that contained relevant control methods related to the built environment.

2.2 Building a knowledge database

At the time of creating the project, whole PDFs could not easily be uploaded or copied pasted directly into commonly available, low-cost LLMs such as the free version of ChatGPT. When uploading large documents, there is a risk of hitting the token limit due to the amount of information the LLM must decipher, therefore making it uneconomical to retrieve data. We therefore needed to create a database from selected PDFs articles that the LLM could easily draw information from. The PDFs were split into smaller 'chunks' of text, the length of which is smaller than the token size that the LLM supports. Each text chunk is then vectorised and given a series of embeddings that measure the relatedness of the text. These embeddings enable easier searching, clustering, anomaly detection and classification of the text (OpenAI Platform 2023). This allows for the smaller embeddings to be compared rather than the larger chunks of text. These embeddings are combined into a Semantic Index which forms the knowledge base. We used LangChain (Langchain 2024) to connect OpenAI's LLM to our own knowledge base.

A copy of the code for the *Aedes* chatbot is available here:

<https://colab.research.google.com/drive/1tPYQfs9Xw5jsq0tF6hwX1zcoVFVFKWb5?usp=sharing>.

2.3 Creating queries for the *Aedes* chatbot

Figure 4 describes the process of information retrieval from the article database and how this is used to generate a response to a given query. In order for the user to ask a question to the chatbot, an embedding is made based on the question's text. A semantic search is then carried out on the knowledge base from the question embedding. A series of ranked results are retrieved from the knowledge base, correlating to the relatedness to the query. These results are then interpreted by the LLM to generate a comprehensible response that is fed back to the user as an answer to their initial question.

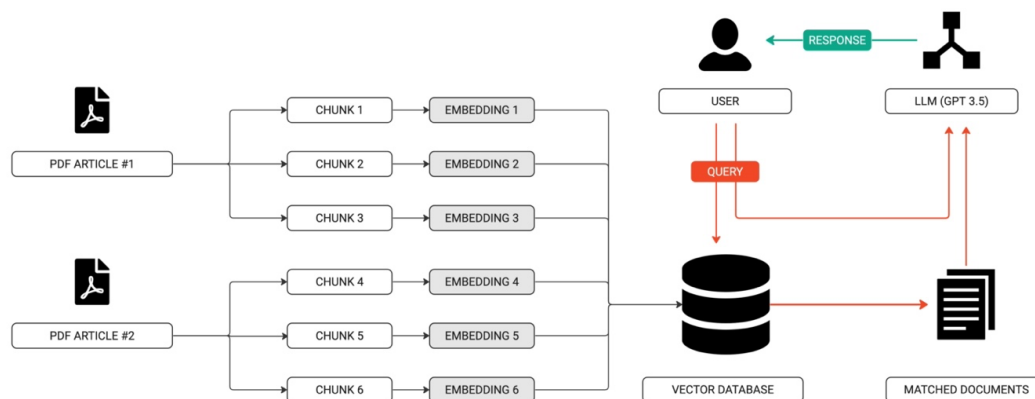


Figure 4: Process of information retrieval, showing how PDFs are separated into smaller chunks that are then embedded and stored in a vector data base. When a query is raised by the used, this vector database is searched to find which chunk is likely to contain relevant information. Those chunks are then fed to the LLM along with the query that is used as the basis of the response. Source: (Author 2023)

2.4 Testing the prototype *Aedes* chatbot

A first working prototype chatbot using python was developed in Google Colab (Google 2017), shown in Figure 5. For initial testing of the application, 11 articles were chosen at random from a previous literature review of *Aedes* vector control interventions and stored in an online Google Drive root folder which the language model can access. UnstructuredPDFloader (Langchain 2024b) was used to read the PDF files which were then divided into chunks based on length and vectorised to create a database of embeddings for the LLM to draw on using LangChain. The model was connected to the OpenAI GPT 3.5 LLM (OpenAI 2022) with a unique OpenAI key. The chatbot was designed to provide simple recommendations on how to alter or modified the built environment to reduce the risk of contracting an *Aedes*-borne disease, such as dengue, based on the literature contained in the knowledge database. Additional code was added to provide the reference of the source paper that the answer was taken from. The quality of the responses was assessed by asking the chatbot a series of simple questions and follow up questions, prepared and reviewed by the authors, a qualified architect and computer scientist. We identified issues with the chatbot's responses and identified strategies to improve the quality of the answers for a future version of the chatbot.

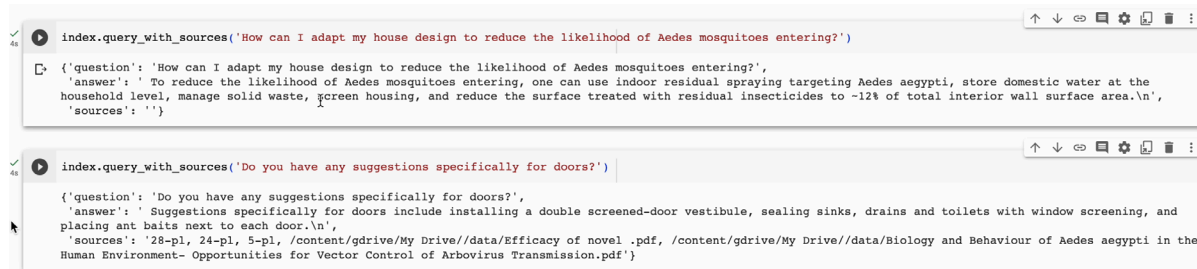


Figure 5: Image of chatbot interface developed in Google Colab, showing a query entry and response. Source: (Author 2023)

3.0 RESULTS

Table 1: Sample questions and responses from the Aedes LLM chatbot. Source: (Author 2023)

Query	Response
“How can I adapt my house design to reduce the likelihood of Aedes mosquitoes entering?”	“To reduce the likelihood of Aedes mosquitoes entering, one can use indoor residual spraying targeting Aedes aegypti, store domestic water at household level, manage solid waste, screen housing, and reduce the surface treated with residual insecticides to ~12% of the interior wall surface area.”
“Do you have any suggestions specifically for doors?”	“Suggestions specifically for doors include installing a double screened vestibule, sealing sinks, drains and toilets with window screening, and placing anti baits next to each door.”
“What plants could I use to reduce mosquitoes?”	“Plants such as O.gratissimum, moringa oleifera and fungi have been reported to be used as repellants against mosquitoes.”

Example questions and responses are given in Table 1. A qualitative assessment of the responses found that for most queries given, an appropriate response was returned that was both relevant to the built environment and based on information from within the scientific literature, however there were some limitations to the answers received. We outlined the limitations and issues identified in Table 2. These included: a limited variability in the responses, with different questions receiving the same or similar answers, inclusion of irrelevant information, the cost associated with making each query, source biases and limited length of the responses. For each of these limitations we were able to identify strategies that could improve the quality of the query responses. These improvements included: increasing the number of articles in the knowledge database, optimising the LLM’s response persona, improved PDF chunking and exploring the use of Opensource LLM’s, such as LLaMa, Alpaca or GPT4All.

Table 2: Summary of issues identified during testing of the Aedes LLM chatbot, and potential areas of improvement. Source: (Author 2023)

Issues Identified		Strategies for Improvement	
1	Responses are relatively limited in variety, in some cases different questions received the same or very similar answers.	1, 2, 4, 5	Increase the number of articles included in the knowledge database.
2	Some responses include information that is still related to Aedes mosquito prevention, however, has little relevance to the query, for example describing drain modifications to a query concerning doors.	4, 5	Modify the ‘persona’ of how the LLM responds. For example, the model can be programmed to ask with the persona of a researcher to obtain a more academic response.
3	There is a cost associated with every query due to the OpenAI being a paid LLM. The test cost was approximately \$0.1-0.3 per query with the test database.	3	Utilise an opensource language model such as LLaMa, Alpaca or GPT4All
4	Preference for certain sources over others in the test database.	1, 4, 5	Create text chunks that have overlap between the paragraphs to improve information retrieval.
5	Responses are often short in length, around one sentence, lack detail.	1, 2, 4, 5	Optimise how the PDF text is ‘chunked’, i.e. if it split by paragraph compared with divided based on text length then information is less likely to be split across multiple ‘chunks’.

4.0 DISCUSSION

Given the rapid advancements in the field of artificial intelligence (AI) and the significant investment required to develop well-functioning large language model (LLM) applications, the Aedes Assistant chatbot was created as a proof-of-concept to investigate the challenges and potential benefits of integrating these technologies into cross-disciplinary architectural projects. The assessment of the chatbot demonstrated its potential as an advisory tool for building environment professionals during the early stages of the design process, particularly in contexts with a high prevalence of Aedes-borne diseases. The chatbot was able to identify relevant control strategies and apply them to the design of the built environment. However, there were still several areas where the quality of the responses could be improved.

The ability to create a custom knowledge database presents various opportunities to ensure oversight over the quality of the source information and carry out tasks on large datasets that would otherwise be very time-consuming. For instance, a landscape architect seeking to identify local plants that mask or absorb mosquito attractants like CO₂ could provide the chatbot with botanical studies and text on the local flora in their context. The chatbot could then quickly pinpoint the plants that are both local and meet the requirements to absorb attractants, a task that would be labor-intensive if done manually.

One limitation of the text-based chatbot format is that the responses will inherently be descriptive rather than visual. In its current capacity, the chatbot can provide text information on different design approaches, which can then be interpreted and developed further by building professionals using the interface. To address this limitation, the chatbot could be integrated with a text-to-image generator, such as DALL-E 3 or Midjourney, to create visual representations of the generated responses.

The financial cost of using OpenAI's LLM presents a barrier to the adoption of such technologies. Open source LLMs, such as LLaMa, Alpaca or GPT4All, offer a promising alternative. This is particularly relevant for Aedes-borne diseases which disproportionately affect Low- and Middle-Income countries, and therefore free and open-source tools could allow great access and impact. There are also access barriers to the articles needed to build a robust database for LLMs to drawing information from, such as journal paywalls and the cost to research teams and scholars publishing open access articles. Given the pace of change within the field of AI and the continuous updating required, these open-source models are reliant on a strong community of unpaid programmers to maintain them.

The use of LLMs for interdisciplinary projects would still require extensive oversight of source information and responses to ensure that the included research meets rigorous academic standards. It would also require a thorough understanding of how these answers are generated and their interpretability, which will become increasingly difficult as the complexity of these models increases over time. LLMs are still largely in their infancy, and there are numerous resource, socio-economic and political challenges these tools will bring, such as the environmental impact of the energy required to compute each query, ensuring job security for previously human done tasks replaced by AI, and how to navigate intellectual property laws for algorithm training when these models have access to the entire internet (Thirunavukarasu et al. 2023).

CONCLUSION

The development of the Aedes Assistant chatbot highlights the potential of customisable LLM applications to assist building professionals working on interdisciplinary projects by helping them gather relevant information, particularly in the early design stages, from fields outside their own expertise. These tools can help facilitate greater understanding and collaboration between disciplines, enabling the synthesis of knowledge and perspectives from various fields. This interdisciplinary approach is crucial for tackling complex issues, such as the increasing health burden of Aedes-borne diseases, which require coordinated efforts across sectors. However, there are also significant challenges with LLM tools. As these generative AI technologies rapidly advance, it will be essential for users to maintain critical oversight of the source data and responses to ensure their interpretability, validity, and applicability. This will require careful evaluation of the data quality, biases, and limitations inherent in the LLM's knowledge base and staying up to date with developments within the field of AI.

REFERENCES

- Achee, Nicole L., Fred Gould, T. Alex Perkins, Robert C. Reiner Jr., Amy C. Morrison, Scott A. Ritchie, Duane J. Gubler, Remy Teyssou, and Thomas W. Scott. 2015. "A Critical Assessment of Vector Control for Dengue Prevention." *PLOS Neglected Tropical Diseases* 9 (5): e0003655. <https://doi.org/10.1371/journal.pntd.0003655>.
- Carrasco-Tenezaca, Majo, Musa Jawara, Mahamed Y. Abdi, John Bradley, Otis Sloan Brittain, Sainey Ceesay, Umberto D'Alessandro, et al. 2021. "The Relationship between House Height and Mosquito House Entry: An Experimental Study in Rural Gambia." *Journal of the Royal Society, Interface* 18 (178): 20210256. <https://doi.org/10.1098/rsif.2021.0256>.
- Google. 2017. "Google Colaboratory". <https://colab.research.google.com/>.
- Kasneji, Enkelejda, Kathrin Seßler, Stefan Kuechemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, et al. 2023. "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education." *Learning and Individual Differences* 103 (January):102274. <https://doi.org/10.1016/j.lindif.2023.102274>.
- Langchain. 2024a. "Langchain". <https://www.langchain.com/>.
- _____. 2024b. "UnstructuredPDFloader". https://api.python.langchain.com/en/latest/document_loaders/langchain_community.document_loaders.pdf_unstructuredpdfloader.html.

- Mshamu, Salum, Arnold Mmbando, Judith Meta, John Bradley, Thomas Bøjstrup, Nicholas Day, Mavuto Mukaka, et al. 2022. "Assessing the Impact of a Novel House Design on the Incidence of Malaria in Children in Rural Africa: Study Protocol for a Household-Cluster Randomized Controlled Superiority Trial." *Trials* 23 (June). <https://doi.org/10.1186/s13063-022-06461-z>.
- Murray, Kris A., Luis E. Escobar, Rachel Lowe, Joacim Rocklöv, Jan C. Semenza, and Nick Watts. 2020. "Tracking Infectious Diseases in a Warming World". *BMJ* 371: m3086. <https://doi.org/10.1136/bmj.m3086>.
- Obonyo, Esther, Sumit Pareek, and Dawit O. Woldu. 2019. "Decision Making within the Built Environment as a Strategy for Mitigating the Risk of Malaria and Other Vector-Borne Diseases." *Buildings* 9 (1). <https://doi.org/10.3390/buildings9010002>.
- Oliveira, Sandra, Jorge Rocha, Carla A. Sousa, and César Capinha. 2021. "Wide and Increasing Suitability for *Aedes albopictus* in Europe Is Congruent across Distribution Models." *Scientific Reports* 11 (1): 9916. <https://doi.org/10.1038/s41598-021-89096-5>.
- OpenAI. 2022. "GPT-3.5." "OpenAI Platform". 2023. <https://platform.openai.com>.
- Scao, Teven Le, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, et al. 2022. "BLOOM: A 176B-Parameter Open-Access Multilingual Language Model". *ArXiv* abs/2211.05100. <https://api.semanticscholar.org/CorpusID:253420279>.
- Seidlein, Lorenz von, Konstantin Ikonomidis, Salum Mshamu, Theresia E. Nkya, Mavuto Mukaka, Christopher Pell, Steven W. Lindsay, Jacqueline L. Deen, William N. Kisinza, and Jakob B. Knudsen. 2017. "Affordable House Designs to Improve Health in Rural Africa: A Field Study from Northeastern Tanzania." *The Lancet Planetary Health* 1 (5): e188–99. [https://doi.org/10.1016/S2542-5196\(17\)30078-5](https://doi.org/10.1016/S2542-5196(17)30078-5).
- Shragai, Talya, Blanka Tesla, Courtney Murdock, and Laura C. Harrington. 2017. "Zika and Chikungunya: Mosquito-Borne Viruses in a Changing World." *Annals of the New York Academy of Sciences* 1399 (1): 61-77. <https://doi.org/10.1111/nyas.13306>.
- Thirunavukarasu, Arun, Darren Ting, Kabilan Elangovan, Laura Gutierrez Sinisterra, Ting Tan, and Daniel Ting. 2023. "Large Language Models in Medicine." *Nature Medicine* 29 (July). <https://doi.org/10.1038/s41591-023-02448-8>.
- Watts, Nick, Markus Amann, Nigel Arnell, Sonja Ayeb-Karlsson, Jessica Beagley, Kristine Belesova, Maxwell Boykoff, et al. 2021. "The 2020 Report of The Lancet Countdown on Health and Climate Change: Responding to Converging Crises." *The Lancet* 397 (10269): 129–70. [https://doi.org/10.1016/S0140-6736\(20\)32290-X](https://doi.org/10.1016/S0140-6736(20)32290-X).
- World Health Organization (WHO). 2017. "Global Vector Control Response 2017–2030: A Strategic Approach to Tackle Vector-Borne Diseases." Geneva: World Health Organization.
- World Health Organisation (WHO). n.d. "Dengue and Severe Dengue." Accessed 16 October, 2023. <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>.
- Yang, Xi, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E. Smith, Christopher Parisien, Colin Compas, et al. 2022. "A Large Language Model for Electronic Health Records." *npj Digital Medicine* 5 (1): 194. <https://doi.org/10.1038/s41746-022-00742-2>.

