January 28, 2025 (revised)
Human Nature Lab
Large Language Model (LLM) and Generative AI Use Policy

Although the use of AI in the conduct of science is growing and evolving, the current rules regarding the use of LLMs in our lab are as follows.

Permitted uses

- You may use AI systems (developed in the USA or Europe) to help write and debug analytic code. However, you must personally review every single line of code and be able to vouch for it. You can use AI agents as assistants and as debuggers, but not to replace your own agency and responsibility.

- You may use AI to help conceive of appropriate graphical images (e.g., for cover proposals) or to edit images (e.g., certain sorts of non-data-containing aspects of figures and slides, such as illustrative cartoons, etc.).

- You may use AI to help perfect ideas (e.g., "I'm thinking of analyzing how X affects Y using procedure Z; are there other aspects of this I should think about?" or "I think claim X must be true; what are some possible reasons this might be or might not be the case?"). Relatedly, you may use it to help you learn new things (e.g., "Give me a summary of topic X at Y level of complexity").

Prohibited uses

- Do not use AI to write prose for any of our manuscripts, in any way. This is for several reasons: (1) concern about hallucinations (though this problem is getting better rapidly); (2) concern for the risk of us inadvertently engaging in plagiarism (e.g., see this paper: https://arxiv.org/abs/2409.12367); and (3) likelihood of being detected. Regarding the last point, when we submit papers, most journals now ask that we certify that we have not used LLMs for writing. Lying about such a thing is obviously unacceptable. Moreover, the increased availability of watermarking technology is making it easier for journals to enforce this policy. (e.g., https://www.nature.com/articles/s41586-024-08025-4). Relatedly, we should not use LLMs to help edit manuscripts. The reason is that inputting manuscript text for editing or clean-up before submission could inadvertently lead to leaks of our unpublished work! Someone else using the same LLM after we used it might get the benefit of our prose or even findings, for example.

- It is not OK to ever upload actual data from our lab into AI systems (e.g., uploading a file and asking the AI agent to detect patterns or plot results). A primary concern here is that our data (e.g., from Honduras) is often confidential and there is no assurance that the LLM provider will not keep a copy of our data to perfect its models. More generally, it is longstanding policy of HNL that no data may be migrated off Andromache without my, Tom's, or Liza's written permission (an exception is data that we already have released

publicly, as we sometimes do with our breadboard experiments, or stripped and anonymized data that we do sometimes migrate).

- In particular, it is not permissible to upload any lab-related materials (grants, papers, data, output, images, etc.) to DeepSeek or other Chinese LLMs. The primary version of DeepSeek keeps all uploaded materials to further train its models, and (while this is also true of OpenAI and many other LLMs), in the case of DeepSeek, the Chinese Communist Party has access to any such materials. This creates numerous individual and collective hazards. There are versions that can be personalized to avoid this hazard; in that case, please discuss any plans to use such versions of DeepSeek with me in writing before sharing any lab materials with such derivative models.