ARTIFICIAL INTELLIGENCE FOR AB INITIO ENERGIES: THE "AI ENERGIES" DATABASE

NIKE DATTANI[a], *Computer Spectroscopy Group, HPQC Labs, Waterloo, Canada*.

The AI ENERGIES database was first made available on GitHub in January 2018, and by February 2023 has had more than 600 commits involving *ab initio* ground and excited state energies for various atoms and molecules with various charges, spin multiplicities and geometries. It attempts to make available all *ab initio* calculations ever done by participants, and to use artificial intelligence to predict energies for systems that are too difficult for running full *ab initio* calculations.

What began as a repository to preserve output files and summaries of various calculated energies for the Li_2 molecule, and to make them available for artificial intelligence training, has grown now to include molecules as big as $C_{12}H_{10}B_2N_2$, and state-of-the-art benchmark calculations on systems as small as the H atom but with basis sets as large as aug-cc-pV10Z.

The repository also contains an FCI database which attempts to curate and make available all full configuration interaction energies known, and it contains a coupled cluster (CC) database which attempts to include an unusually thorough level of detail about all coupled cluster calculations ever done by participants. By including the maximum RAM and CPU time used, along with as many details about the system for which an FCI or CC calculation is performed, and details about the hardware used, estimations can be made in advance for the cost (or feasibility) of doing for example a CCSDTQ(P) calculation on a singlet C_{2v} system with 8 correlated electrons in 200 spatial orbitals with three CPU threads: the estimation in this case would be approximately two weeks with 147GB of RAM. In this way, scientists can refer to the database and the predictions that it is able to provide, which can help with the planning of projects and the evaluation of a project's feasibility.

The repository additionally contains a `GENBAS` file which contains 70,000+ lines of Gaussian basis set exponents and contraction coefficients with an unusually large amount of care towards reproducibility and inclusion of some of the most coveted and specialized basis sets that are not available in the ordinary basis set repositories available online.

Anyone is able to contribute data to the AI ENERGIES database on GitHub (https://github.com/HPQC-LABS/AI_ENERGIES), and this can help them preserve and organize data that might otherwise get accidentally deleted or lost. The database also has a Digital Object Identifier (DOI): https://doi.org/10.5281/zenodo.5529103

[a]nike@hpqc.org