**EleutherAI (GPT-NeoX) - total: 34/48**

1. Data sources
   * 4 points: Dataset pipeline is disclosed as a singular dataset derived from The Pile with all relevant preprocessing, with The Pile itself being documented extensively (including dataset size and fine-grained sourcing) in the associated paper and datasheet.
   * Sources: <https://huggingface.co/EleutherAI/gpt-neox-20b#training-dataset> and <https://arxiv.org/abs/2201.07311>
2. Data governance
   * 3 points: The subject of data governance and curation is raised, with specific practices to mitigate harms (e.g. filtration practices). With that said, no constraint is imposed to formally preclude the use of inappropriately-governed data, though many dimensions of governance are considered.
   * Source: <https://arxiv.org/pdf/2201.07311.pdf>
3. Copyrighted data
   * 4 points: Copyright is explicitly discussed, highlighting “Some of the documents in the datasets that the Pile is based on are copyrighted.In particular, Books3 is almost entirely comprised of copyrighted works, and a substantial portion of arXiv and PhilPapers are as well. Other datasets, such asPubMedCentral and GitHub contain documents that may be under limited licensing, but are not copyrighted as far as we are aware.”
   * Source: <https://arxiv.org/pdf/2201.07311.pdf>
4. Compute (additive)
   * 4 points: The model size, training time, hardware, and FLOPs are disclosed.
   * Source: <https://arxiv.org/pdf/2204.06745.pdf>
5. Energy (additive)
   * 4 points: The energy usage, emissions, energy sources, and measurement methodology are clear. No extensive discussion is provided of energy usage mitigation, though various decisions are taken to improve efficiency (e.g. use of Megatron and DeepSpeed libraries, parallelism strategy) that do contribute to reduced energy use.
   * Source: <https://arxiv.org/pdf/2204.06745.pdf>
6. Capabilities and limitations
   * 4 points: Capabilities are described (“GPT-NeoX-20B was developed primarily for research purposes. It learns an inner representation of the English language that can be used to extract features useful for downstream tasks”), with several acknowledged limitations (e.g. “not been fine-tuned for downstream tasks for which language models are commonly deployed, such as writing genre prose, or commercial chatbots. This means GPT-NeoX-20B will likely not respond to a given prompt the way products such as ChatGPT do”, “English-language only, and thus cannot be used for translation or generating text in other languages”, “Never rely on GPT-NeoX-20B to produce factually accurate output.”), largely aligning with the literature on language model limitations. Discussion of fine-tuning is specific/more clearly grounded to this model than all models in general.
   * Source: <https://huggingface.co/EleutherAI/gpt-neox-20b>
7. Risks and mitigations (additive)
   * 1 points: Risks are listed (e.g. bias, factual incorrectness), though no mitigations are discussed.
   * Source: <https://arxiv.org/pdf/2204.06745.pdf> and <https://huggingface.co/EleutherAI/gpt-neox-20b#limitations-and-biases>
8. Evaluations (additive)
   * 1 point: Extensive evaluation for accuracy using EleutherAI LM Harness, though no evaluations for unintentional (e.g. bias) or malicious (e.g. disinformation) harms, as well as other desiderata (e.g. robustness, calibration).
   * Source: <https://arxiv.org/pdf/2204.06745.pdf>
9. Testing (additive)
   * 0 points: No specific discussion of further internal testing, external evaluations, or external redteaming.
   * Source: <https://arxiv.org/pdf/2204.06745.pdf>
10. Machine-generated content
    * 3 points: The documentation indicates “Please inform your audience that you are using artificially generated text.”, though there is no clear enforcement mechanism for downstream use.
    * Source: <https://huggingface.co/EleutherAI/gpt-neox-20b#limitations-and-biases>
11. Member states
    * 2 points: No known country-level restrictions on use for weights hosted on <https://the-eye.eu/>. No discussion of how the models are deployed across different EU member states.
    * Sources: <https://the-eye.eu/>
12. Downstream documentation
    * 4 points: Documentation of model and data, with documentation accompanying model both on the Hugging Face page and the <https://the-eye.eu/> download, as well as instructions on appropriate use subject to documentation.
    * Source: <https://huggingface.co/EleutherAI/gpt-neox-20b> and <https://the-eye.eu/public/AI/models/GPT-NeoX-20B/20B_model_card.md>