Can Private LLM Agents Synthesize Household Energy Consumption Data?

Mahathir Almashor*
mahathir.almashor@csiro.au
CSIRO Energy
Sydney, Australia

Yusuke Miyashita*
yusuke.miyashita@csiro.au
CSIRO Energy
Melbourne, Australia

Sam West*
sam.west@csiro.au
CSIRO Energy
Newcastle, Australia

Thi Van Dai Dong*
thivandai.dong@csiro.au
CSIRO Energy
Wollonggong, Australia

ABSTRACT

Reproducible science requires easy access to data, especially with the rise of data-driven and increasingly complex models used within energy research. Too often however, the data to reconstruct and verify purported solutions in publications is hidden due to some combination of commercial, legal, and sensitivity issues. This early work presents our initial efforts to leverage the recent advancements in Large Language Models (LLMs) to create usable and shareable energy datasets. In particular, we're utilising their mimicry of human behaviors, with the goal of extracting and exploring synthetic energy data through the simulation of LLM agents capable of interacting with and executing actions in controlled environments. We also analyse and visualise publicly available data in an attempt to create realistic but not quite exact copies of the originals. Our early results show some promise, with outputs that resemble the twin peak curves for household energy consumption. The hope is that our generalised approach can be used to easily replicate usable and realistic copies of otherwise secret or sensitive data.

CCS CONCEPTS

• **Computing methodologies** → *Multi-agent systems*; **Natural language generation**; • **Security and privacy** → **Social aspects of security and privacy**; • **Information systems** → **Data analytics**.

KEYWORDS

Synthetic Data, Generative AI, Large Language Models, Household Electricity Consumption

ACM Reference Format:
Mahathir Almashor, Yusuke Miyashita, Sam West, and Thi Van Dai Dong. 2024. Can Private LLM Agents Synthesize Household Energy Consumption Data?. In *The 15th ACM International Conference on Future and Sustainable Energy Systems (E-Energy '24), June 04–07, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3632775.3661993

*Commonwealth Science and Industrial Research Organisation (CSIRO) Energy

E-Energy '24, June 04–07, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0480-2/24/06
https://doi.org/10.1145/3632775.3661993

1 INTRODUCTION

The energy community is working tirelessly on the transition to renewables and there is a pressing need to share data across research organisations. However, the sharing of useful and contemporary data in many domains is often an exercise fraught with hurdles and competing motives. The barriers to effective sharing range from privacy and cyber-security concerns, through to competing commercial interests. Given the urgency and enormity of the task at hand, we need a way for partners and collaborators across the globe to effectively evaluate their solutions against agreed datasets.

It is in this spirit that we present this very early work exploring the generation of synthetic yet realistic data within the energy domain. Our aim is arrive at method to generate raw household energy consumption data that (i) reflects the demographics and behaviours of a particular geographic region; (ii) be realistic and emergent in a way that accounts for the seemingly stochastic nature of human actions; and (iii) be free of the encumbrances that typically prevent the wider dissemination of data between institutions.

To this end, we propose the use of private Large Language Models (LLMs) [3] in the synthesis of such data. We see a range of possibilities in the abilities of current models to drive interactions within multi-agent simulations, and want to leverage their unique so-called hallucinations to arrive at realistic behavioural patterns. Put simply, we want our LLM-powered agents to "dream" about their day-to-day actions and organically arrive at energy consumption patterns that mirror real-life. The reasons are as follows:

- As was seen in [11], emergent behaviours were seen when the authors simulated a small town of 25 agents powered by ChatGPT [10]. This included an unscripted event where the agents created a mayoral election and then proceeded to realistically interact with one another about it. This is the same phenomena we wish to capture, where the agents can naturally vary their behaviours and activities during the day to form a better mimicry of real-life humans. This includes variations that account for their specific characterisations (i.e., identity), occupations, and interactions with immediate family members.
- We use the term Private LLM both in terms of their localised nature, as well as their inherent privacy benefits. That is, rather than depend on cloud-based and costly implementations such as

OpenAI's ChatGPT, we strove for smaller models that can be run within any organisation's infrastructure. While the performance of these local models may not be on par with their more famous brethren, better accessibility and lower costs presents a good counterbalance. In this way, the barrier to entry is lowered for most institutions when replicating our methods and results.
- Private also refers to the ability to keep an organisation's valuable intellectual property safer and only within its confines. There is no need to share text prompts, techniques and potential seed data with any third-party service as everything is run on-premise or within that organisation's own cloud infrastructure.

There is another dimension which delves further into privacy, and is the reason why we chose household energy consumption at the first attempt. Simply put, it is difficult to obtain current, usable and customisable datasets because it invariably impacts real-life privacy concerns. We cannot effectively measure the daily consumption of a human household, let alone the hundreds or thousands within a geographical region. In commercial settings, there are legal constraints on the collection of consumption data [13], given the intrusiveness and sensitivities such gathering would entail. We often have to rely on observations, household demographics, presence of appliances, and wider energy usage that are entirely self-reported [1], with all the data quality issues that entails. In any case, electrical load is typically only measured at the meter, requiring heuristics or additional hardware to ascertain the presence and consumption of individual appliances. It is impractical to measure consumption of each appliance and outlet for an entire city, not to mention how invasive and time-consuming this would be for any given household. Our contributions may be summarised as follows:

- A novel addition to an LLM-enabled simulation engine to generate emergent daily routines of multiple agents, and the subsequent extraction of corresponding energy patterns.
- The customisation of an existing simulation engine with a private LLM implementation called Mistral that can be run within localised infrastructure and without costly ongoing access fees.
- Experimentation with our approach that exhibits promising ability to replicate the consumption patterns seen in publicly available datasets that were discovered and analysed.

Ethical Considerations

No personally identifiable information (PII) or other sensitivities were found in the household electricity consumption data used beyond anonymised age-brackets, demographics and post-codes.

2 BACKGROUND & RELATED WORK

2.1 Synthetic Energy Data

The use of GANs for generating energy time series data has recently gained prominence, fueled by the increased use of time series data across various domains. One objective in generating time series is to accurately capture temporal dynamics. Some earlier approaches, such as C-RNN-GAN [9] and RGAN/RCGAN [4] have been developed to learn the temporal variations of data. However, these approaches require real data for model train, which in turn poses the risk of leaking sensitive information in any generated outputs.

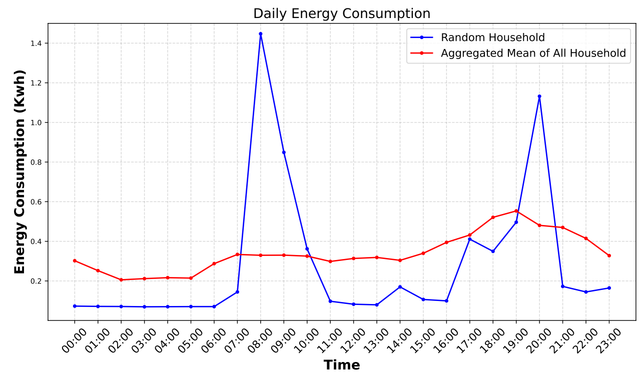

Figure 1: Snapshots of daily and aggregated mean of typical household energy consumption.

Our approach attempts to overcome this issue by conducting household activities within an LLM-powered simulation world, and subsequently extracting energy data from those activities. This essentially avoids the use of any real data altogether and thus minimising any privacy or sensitivity concerns. In other words, there is little to no risk of inadvertently replicating a household's exact consumption patterns, as the genesis of the data creation is entirely independent from any real-life source.

2.2 LLM and LLM Agents

Large Language Models, epitomised by groundbreaking models like OpenAI's GPT-3[2], FaceBook's Llama2[14] and Mistral[6] AI's eponymous model have showcased strong abilities to interpret, generate, and simulate human-like text. It would be an understatement to say that LLM technologies have gained popularity recently, with various research looking at everything from teasing out hidden meanings in speech [5] to using their generation capabilities in software programming tasks [7].

Beyond the obvious chat-bot category of applications, there is also a significant amount of research in LLM agents aimed at imitating human behavior. A primary example of this is Simulacra [11], which introduces a fusion between LLM and computationally interactive agents in a sandbox environment to enable believable simulations of human behavior. Similarly, MemGPT [8] uses LLM as an operating system, allowing it to think, reflect, and produce actions to interact with external devices. Here, we utilized the Mistral-7B private LLM to power the Simulacra agents instead of the original ChatGPT backend, due to cost and accessibility reasons.

2.3 Benchmark Datasets

The Smart Grid Smart City Customer (SGSC) [1] data was collected between 2010 and 2014 as part of a joint industry and government initiative. It was one of only a few openly available datasets on household energy consumption that we discovered. It contains 30-minute interval readings of electricity usage and generation (measured in kWh) for 78,720 participating customers, of which we only had access to a subset of 13,735 households.

In addition, we also discovered the Solar Cities dataset [12], which contained energy consumption and generation information for almost 38,000 homes in seven Australian cities, recorded at 30-minute intervals from 2005 to 2013. After an extensive pre-processing step, we focused on only 4,332 households due to various data quality and accessibility issues. This data was collected mainly as part of a governmental initiative design to measure the impact of direct interventions on household usage patterns.

In much the same way, there is the possibility to design for and simulate interventions to consumption patterns within our proposed approach. In [12], consumption was measured for households before and after an intervention such as the installation of solar panels, in an attempt to gauge their impacts to electricity demands. Similarly, we can use our LLM-enabled approach to simulate the same objective, by introducing interventions and appropriate pricing dynamics, and watching how the agents respond. There is also the ability to model the impact of incentives on the agents, and measure the organic spread of interventions introduced gradually.

Figure 1 shows the daily energy consumption for randomly chosen household (in blue) as well as aggregated mean of all households in our dataset (in red). Daily energy consumption varies for each household, but we see the trend of two peaks in the morning around 8am and evening around 7pm. We see the typical morning and evening peaks of energy usage in the aggregated mean, which is a well-known phenomena observed in individual households and up to the wider grid. On the other hand we witness the expected variability in energy usage for a single household on any given day.

Figure 2: Example simulation step of two agents in their daily activities, with one using an electrical appliance (their TV).

3 APPROACH

The approach involves two stages, with the first stage consisting of running the simulacra, and the second involving the extraction of household energy data from the simulation using a variety of methods to perform this extraction. For each step of the simulation, description of each persona's actions and objects they are interacting with is generated. They are in the format of "Persona A is doing Action B at Location C". From these, appliances which consumes energy are identified. The advantage of this two-stage approach is its applicability onto any other LLM simulator. That is, simply allow the LLM to record its actions at each step, and then proceed with extracting useful information from it.

3.1 Private LLM in Simulacra

Utilizing ChatGPT or any other API-based LLM can offer convenience, reliability, and strong performance. However, this approach may entail transferring sensitive data and routing it through external services. In contrast, the deployment of Private/Local LLM enables the execution of the entire simulation within closed environments, ensuring the security and privacy of the data.

We utilized Mistral-7B as it is one of the highest performing smaller models which are relatively fast when running iterative experiments, and more importantly, can be fit into more modest compute infrastructures. "7B" here refers to 7 Billion parameters, with some LLM models reaching 65B and more. Suffice it to say for now that the smaller the number of parameters, the more manageable it is to run within an organisation's infrastructure.

This smaller Mistral model is then plugged into the Simulacra platform [11] as its LLM engine. The authors recommend a robust LLM for the agents, ideally as proficient as or superior to ChatGPT-3.5 Turbo as using smaller models may degrade the simulation quality. However, we've found that they still exhibit the ability to generate reasonable plans and actions, resulting in a plausible human-like energy patterns. For simplicity, we've reused the prompts within the original paper to condition our LLM agents.

3.2 Energy Data Extraction

3.2.1 String Match. In this straightforward approach, we employ string matching to determine whether the descriptions of daily activities from the simulations include energy appliances. This baseline is perhaps the most accurate reflection of energy data extraction from agent activities. For example in Figure 2, the simulation step is described as "Maria Lopez is watching TV@common room sofa, Klaus Mueller is having dinner@kitchen sink", the appliance `TV` is easily matched and included as a proxy for energy use. The identified list of appliances in the simulations are as follows:

`TV`, `shower`, `refrigerator`, `toaster`, `cooking area`, `microphone`, `piano`, `game console`, `computer desk`, and `computer`.

So, for every mention of an appliance in each time-step's description, we mark that as an active use of that appliance. A simple counting of all the appliances used in the household is performed for each time-step in the simulated day. This allows us to get a measure of energy consumption purely from the activities of agents with regards to appliances, as they go about their day.

3.2.2 Other Approaches. In our quest to extract more accurate energy usage, we also experimented with Semantic Embedding, which employs a text encoder to match the description of LLM generated interactions against appliances based on semantic meaning. One advantage with this method is its ability to capture synonyms or phrasings describing the same appliance. For instance, when the description states "Maria is streaming on Twitch", we can surmise that this would involve the use of either a game console or a computer. This semantic embedding method may better capture such cases compared to basic text matching. However, setting a threshold for matching top-K text embeddings was found to be challenging and resulted in inaccuracies with the extraction of energy data.

We also experimented with using the LLM engine itself to infer appliance usage. LLMs have the capability to extract potential appliance use even if they are not present in the environment, or if the LLM agents are not explicitly described as interacting with them. Thus, with a line like "Maria reads a book while listening to

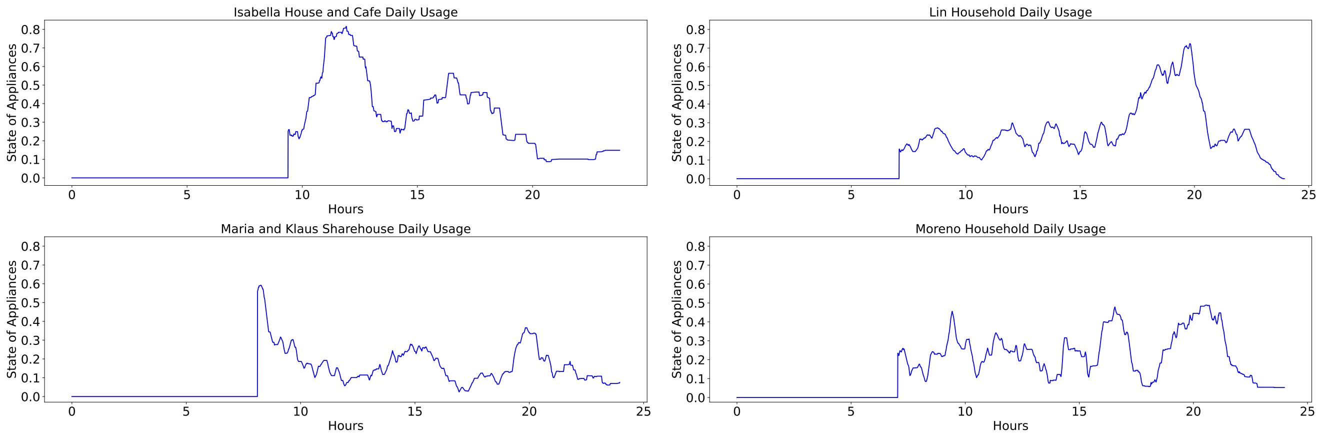

Figure 3: Single day energy usage for four simulated households

music", the LLM may infer that an appliance like a radio or TV is also turned on for the purposes of that background music. However, the challenge here is the occurrence of hallucinations, making it difficult to precisely control their outputs. After extensive analysis, both Semantic and LLM methods were abandoned in favour of the basic but most accurate string matching technique.

4 RESULTS

The status of each appliance at every step is binary, indicating whether it is on or off. This information is consolidated across all appliances to generate discrete values for energy usage. Furthermore, we implemented a rolling mean with a 1-hour window to smooth out our data, aiming to alleviate the instantaneous rises and drops in energy usage, given that energy usage is treated in binary states. Figure 3 depicts the daily energy usage patterns of our four simulated households, which include (clockwise from top-left) Isabella, Lin, Moreno, and Maria and Klaus. This is represented against the rolling mean of appliance states at each time-step.

As can be seen, the daily energy data for each household varies according to their routines, occupations, and lifestyles. This is exactly the variety in energy usage patterns we were aiming for, which resembles the variances seen in recorded single-day usage seen in Figure 1. The energy usage reflects the simulated activities of the LLM agents. For example, the agent Isabella wakes up around 6 am, opens her cafe at 8 am, works at the counter until 8 pm, closes the cafe, and goes to bed around 11 pm. These activities correspond to actions such as turning on the lights in the cafe, serving customers, using the refrigerator, toaster, and cooking area, which collectively contributes to the overall energy consumption pattern.

Similarly, these variances are observed for the Lin and Moreno households. The unique shapes of each household's curve indicates that both the prompt of the LLMs and the simulation environment influence their activities, resulting in their distinct energy consumption curves. This suggests that by conditioning the prompt, one can generate tunable synthetic energy data for the desired household, within the limits of the simulated environment.

Furthermore, to determine the average daily energy usage per person, the mean is calculated from the combined data by summing all the households' consumption at each time-step. In diagram 4, we illustrate two peaks near morning and evening times, capturing the demand periods often seen on weekdays in many aggregated energy datasets. The fluctuations and small peaks during the day may arise from the nature of the simulation environments, such as Isabella running a cafe (which is counted against her household's data) and LLM agents returning home for lunch breaks.

5 CONCLUSION

The proposed approach uses private LLMs to synthesize daily household energy consumption patterns. This is done with privacy in mind, and by using the emergent properties of LLMs to arrive at realistic datasets that can be then be freely shared amongst the energy community. The focus is also on using less computationally intensive and costly LLMs that allows for much easier adoption.

5.1 Limitations & Future Work

The simulation results heavily depend on the capabilities of the LLM, as more advanced LLMs would lead to more realistic simulations and extracted energy data. While early results somewhat resembles the real data, the LLM simulation outputs only binary states of appliances, which doesn't capture continuously varying loads well, and cannot simulate the majority of detailed activities in the cities from which the real data was measured. The translation from state to actual usage remains a potential area for future research.

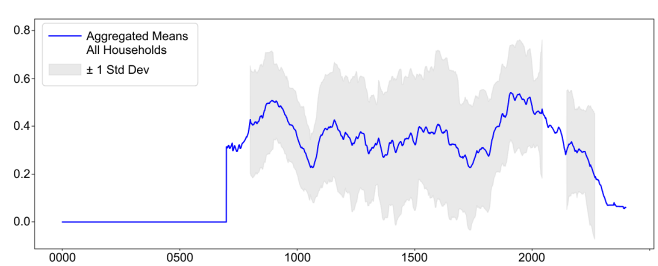

Figure 4: Aggregated mean of energy usage from all four households containing a total of eight agents.

Generating more detailed energy data will require integrating more detailed interactions with appliances, simulating many more agents and will likely require integrating additional factors like climate, weather, traffic, seasons, demographics, and industries. Furthermore, the simulation does not assess climate control systems (e.g. air conditioning), lighting, or transport, despite their significant electricity consumption at home. We posit that they could be treated as constant factors used while occupants are at home, thus not affecting the dynamic usage of appliances. Doubtlessly, there remains a host of further experimentation and rigorous analysis to prove the effectiveness of this approach. The hope is that the wider energy research community will see its value, and collaborate to generate publicly available realistic synthetic datasets.

ACKNOWLEDGMENTS

This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

REFERENCES

[1] Energy Australian Government Department of Climate Change. 2014. Smart-Grid Smart-City Customer Trial Data. https://www.data.gov.au/data/dataset/smart-grid-smart-city-customer-trial-data Last Modified: 2022-04-11T01:46:18.101034.

[2] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. arXiv:2005.14165 [cs.CL]

[3] Vinton G. Cerf. 2023. Large Language Models. *Commun. ACM* 66, 8 (July 2023), 7. https://doi.org/10.1145/3606337

[4] Cristóbal Esteban, Stephanie L Hyland, and Gunnar Rätsch. 2017. Real-valued (medical) time series generation with recurrent conditional gans.

[5] Fan Huang, Haewoon Kwak, and Jisun An. 2023. Is ChatGPT better than Human Annotators? Potential and Limitations of ChatGPT in Explaining Implicit Hate Speech. In *Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 294–297. https://doi.org/10.1145/3543873.3587368

[6] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7B. arXiv:2310.06825 [cs.CL]

[7] Majeed Kazemitabaar, Justin Chow, Carl Ka To Ma, Barbara J. Ericson, David Weintrop, and Tovi Grossman. 2023. Studying the effect of AI Code Generators on Supporting Novice Learners in Introductory Programming. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–23. https://doi.org/10.1145/3544548.3580919

[8] MemGPT Community. 2024. Introduction. https://memgpt.readme.io/docs/index

[9] Olof Mogren. 2016. C-RNN-GAN: Continuous recurrent neural networks with adversarial training.

[10] OpenAI Blog. 2022. Introducing ChatGPT. https://openai.com/blog/chatgpt

[11] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–22. https://doi.org/10.1145/3586183.3606763

[12] Saad Sayeef, Sam West, Stephen Lindsay, Brad Sparkes, and Kate Cavanagh. 2013. Solar Cities Data Analysis Final Report. https://publications.csiro.au/rpr/pub?list=SEA&pid=csiro:EP137924&sb=RECENT&expert=false&n=5&rpp=25&page=1&tr=6&q=Solar%20Cities%20Data%20Analysis&dr=all Publisher: Department of Resources, Energy and Tourism.

[13] René Schwermer, Jonas Buchberger, Ruben Mayer, and Hans-Arno Jacobsen. 2022. Federated office plug-load identification for building management systems. In *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems (e-Energy '22)*. Association for Computing Machinery, New York, NY, USA, 114–126. https://doi.org/10.1145/3538637.3538845

[14] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv:2307.09288 [cs.CL]