Towards Community-Driven Generative AI

Rustem Dautov
0000-0002-0260-6343
SINTEF Digital
Forskningsveien 1
0373 Oslo, Norway
rustem.dautov@sintef.no

Erik Johannes Husom
0000-0002-9325-1604
SINTEF Digital
Forskningsveien 1
0373 Oslo, Norway
erik.johannes.husom@sintef.no

Sagar Sen
0000-0002-5784-7355
SINTEF Digital
Forskningsveien 1
0373 Oslo, Norway
sagar.sen@sintef.no

Hui Song
0000-0002-9748-8086
SINTEF Digital
Forskningsveien 1
0373 Oslo, Norway
hui.song@sintef.no

***Abstract*—While the emerging market of Generative Artificial Intelligence (AI) is increasingly dominated and controlled by the Tech Giants, there is also a growing interest in open-source AI code and models from smaller companies, research organisations and individual users. They often have valuable data that could be used for training, but their computing resources are limited, while data privacy concerns prevent them from sharing this data for public training. A possible solution to overcome these two issues is to utilise the crowd-souring principles and apply federated learning techniques to build a distributed privacy-preserving architecture for training Generative AI. This paper discusses how these two key enablers, together with some other emerging technologies, can be effectively combined to build a community-driven Generative AI ecosystem, allowing even small actors to participate in the training of Generative AI models by securely contributing their training data. The paper also discusses related non-technical issues, such as the role of the community and intellectual property rights, and outlines further research directions associated with AI moderation.**

***Index Terms*—Generative AI, Federated Learning, Crowd-Sourcing, Community, Conceptual Architecture, AI Moderation.**

I. Introduction and Motivation

GENERATIVE Artificial Intelligence (AI) refers to AI models that can generate original content, such as text, images, and music. Unlike traditional AI models that are trained to recognise and classify existing data, Generative AI models learn to generate new data by analysing patterns and structures in large datasets. ChatGPT, developed by OpenAI and sponsored by Microsoft, is admittedly the most prominent example of Generative AI, while similar proprietary services are also developed by the other Big Tech companies.[1]

At the same time, there is a growing interest in open-source AI from smaller companies, research organisations and individual users. It is, admittedly, not feasible for such small players to compete with the Tech Giants individually, but what if they could join forces to collectively challenge the establishing monopoly and lead the development of Generative AI using community-driven democratic principles?

A. Motivating Example: Assisted Code Generation

The fact that these leading tools are proprietarily owned or backed up by big corporations has major implications for their usage and development. One of the biggest concerns is the presence of *bias*, which not only naturally rises from the data used to train the AI and the training algorithm but is also artificially introduced in favour of the corporations' commercial or political interests. This 'intentional' bias can influence consumers who rely on Generative AI tools to make decisions. Another source of bias is filtering, which in theory is supposed to ensure that the generated content meets certain criteria and is appropriate for its intended use. In practice, however, the companies tend to play it safe and apply excessive filtering just to protect themselves from possible ethical scandals. While this is understandable, enforcing such filtering-based moderation may blur important aspects of reality. For example, an AI tool that filters out all mentions of a particular controversial topic may not accurately represent the diversity of opinions.

Another source of bias is that these tools are usually trained only on publicly available data scraped from the Internet, and thus do not account for more specific and nuanced information that is only accessible to private users. For example, if an AI model is trained on public code repositories, it will not incorporate the valuable information exchanged in private corporate networks (*e.g.*, repositories, chats, issue trackers), although these are often considered a more trusted source of professional knowledge than semi-professional answers and informal discussions on Stack Overflow or Reddit. More specifically, in the realm of programming and code generation, these existing models trained on public data sources often overlook a wealth of insights and professional exchanges found in private corporate networks. These networks contain note just code examples, but also specialised practices and innovative solutions, serving as valuable repositories of programming knowledge. Yet, such smaller entities with their unique knowledge are left out from contributing due to privacy and security concerns, and their valuable information is thereby excluded from training. Taken together, these limitations can have significant implications for the accuracy and fairness of the AI-generated output.

[1] Please note that the main focus of this paper is on large language models (LLMs) as the most prominent and representative example of Generative AI, albeit the discussed concepts are applicable to a certain extent to other types of AI-generated content, such as imagery and sound.

B. Paper Contribution and Structure

With this paper we make a first step towards democratising and de-monopolising this emerging market by designing a community-driven Generative AI architecture. The proposed conceptual architecture relies on several existing technologies, which collectively represent a promising toolkit for building a whole open ecosystem for Generative AI. Some key features of the envisioned solution are the ability to preserve data privacy, unbiased and fair model training, decentralised operation, and transparent content moderation, among others. We claim that the emerging field of Generative AI should not be monopolised by the few Tech Giants, but rather collaboratively developed and moderated by an open community of multiple stakeholders, each providing their own perspective on this challenging, yet exciting technology. In explaining the envisioned approach, we also draw parallels with the core elements of democracy to better communicate the proposed concepts and ideas.

The main contribution of this paper is a conceptual architecture of a community-driven ecosystem for Generative AI. The description of this architecture is organised as follows. Section II presents the main technologies underpinning the design of the proposed architecture and describes their roles and benefits. Section III draws parallels with similar relevant projects and critically discusses assumptions and some further research considerations. Section IV summarises the paper with some concluding remarks.

II. Technological Building Blocks

We now present the envisioned conceptual architecture by describing its individual 'building blocks', as depicted in Fig. 1. The architecture can be seen as a vertical stack of technologies, which, we believe, provide a viable foundation for building a community-driven Generative AI ecosystem. The individual layers of the proposed stacked architecture build upon one another, each providing technological foundation for building the next layer. This layered structure is explained in the following subsections starting from the very bottom layer of hardware infrastructures.

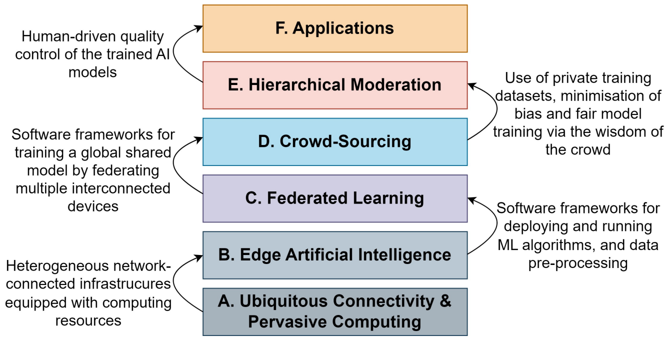
F. Applications
Human-driven quality control of the trained AI models
E. Hierarchical Moderation
Use of private training datasets, minimisation of bias and fair model training via the wisdom of the crowd
D. Crowd-Sourcing
Software frameworks for training a global shared model by federating multiple interconnected devices
C. Federated Learning
Software frameworks for deploying and running ML algorithms, and data pre-processing
B. Edge Artificial Intelligence
Heterogeneous network-connected infrastrucures equipped with computing resources
A. Ubiquitous Connectivity & Pervasive Computing

Fig. 1. Main elements of a community-driven Generative AI ecosystem.

A. Ubiquitous connectivity and pervasive computing

Recent technological advances have paved the way for *ubiquitous connectivity* [1] and *pervasive computing* [2] – the two concepts which revolve around the idea of seamless and pervasive access to computing resources and services. With the pervasive availability of network connections, devices and systems are enabled to be seamlessly connected to the Internet or other communication networks. It emphasises the widespread access to high-speed Internet, wireless networks, and advanced communication technologies. The goal of ubiquitous connectivity is to ensure that people and devices can communicate and access information from anywhere, at any time. This connectivity enables the exchange of data, collaboration, and interaction among various devices and systems.

At the same time, pervasive computing extends the concept of ubiquitous connectivity by focusing on the integration of computing capabilities into everyday objects and environments. It involves embedding intelligence into a wide range of personal devices and human-centred spaces, such as smartphones, 'wearables', household appliances, vehicles, buildings, *etc.*. The goal is to create an environment where computing and information processing become seamlessly integrated into people's daily lives, without requiring explicit user intervention. Together, these technological trends support the growth of emerging technologies such as the Internet of Things (IoT), smart cities, autonomous vehicles, real-time analytics, and other applications requiring low latency, high reliability, and efficient use of network resources [3].

Foundation for the next layer: Ubiquitous connectivity and pervasive computing provide a distributed infrastructure of heterogeneous network-connected devices equipped with computing resources.

B. Edge Artificial Intelligence

The described advances in networking and computing capabilities of field-deployed devices underpin another relevant concept – *edge computing*, which is a decentralised computing paradigm that brings data processing and computation closer to the data source or the 'edge' of the network, instead of relying solely on centralised cloud servers [4]. In edge computing, data processing and analytics are performed at or near the device/sensor level instead of sending all data to a remote data centre for processing. The advantages of edge computing include:

- *Reduced latency*: By processing data locally at the edge, response times can be significantly improved, enabling near-real-time applications.
- *Bandwidth optimisation*: Edge computing reduces the amount of data that needs to be transmitted to the cloud or data centre, thus optimising network bandwidth usage and reducing costs.
- *Enhanced privacy and security*: Local data processing at the edge can help protect sensitive information by reducing the exposure of data in transit and allowing for localised security measures.
- *Offline functionality*: Edge devices can continue to operate and provide services even when connectivity to the cloud is disrupted, ensuring uninterrupted functionality.

Data processing at the edge can range from simple data pre-processing operations to rather advanced AI analytics using Machine Learning (ML). The latter, commonly known as *Edge AI*, refers to the deployment of AI algorithms and models directly on edge devices, such as smartphones, IoT devices, edge servers, and other similar computing nodes [5]. It brings AI capabilities and decision-making closer to the data source, minimising the need for data transmission to centralised cloud servers. The key idea behind Edge AI is to run complex ML-driven data analytics locally at the edge device itself, without relying on continuous cloud connectivity or sending data to remote data centres. This enables near-real-time inference, reduces latency, saves bandwidth, enhances privacy, and enables offline functionality even in the absence of the Internet connection. All these features are especially important to the healthcare domain where physiological data collected by wearable or portable medical devices are processed either directly on those devices or on a smartphone acting as a wireless gateway [6], [7]. Similarly, the data privacy and network bandwidth constraints are usually critical aspects in various image and video recognition scenarios involving CCTV cameras [8], [9].

Foundation for the next layer: Edge AI provides software frameworks for deploying and running ML algorithms, as well as data pre-processing on top of heterogeneous, potentially resource-constrained edge hardware infrastructures.

C. Federated Learning

A natural next step in the Edge AI development was not only to deploy pre-trained AI models and run inference, but also to train the models at the edge. While individual edge devices are still constrained in their computing capabilities to perform heavy-weight model training, the promising solution was to combine multiple devices into an aggregated pool of computing resources and then orchestrate the iterative model training process. This ML approach, known as *federated learning*, enables training models on decentralised data without the need to transfer raw data to a central server [10]. In federated learning, the training process takes place directly on edge devices, such as smartphones, IoT devices, or local servers, where the data is generated and stored. The main idea is to bring the model training to the data rather than to move the data to a central location.

Some prominent federated learning frameworks actively developed and used by the community include Flower,[2] Tensorflow Federated,[3] and OpenFL.[4] Federated learning has applications in various domains, including healthcare, finance, smart devices, and more. It allows for collaborative learning while maintaining data privacy, making it a promising approach for training models on sensitive or distributed data sources. In the context of Generative AI, federated learning can be applied to train LLMs in a distributed and privacy-preserving manner. Instead of centralising the training data on a single server, training can be performed directly on edge devices or local servers where the data resides. The main benefits of federated learning applied to Generative AI include:

- *Privacy*: Federated learning preserves data privacy since the raw data remains on the edge devices and is not directly shared with the central server. This is particularly important when dealing with sensitive user data.
- *Data localisation*: Federated learning enables training on data that is distributed across multiple devices or locations, allowing for localised training and personalised models while avoiding data silos.
- *Efficiency*: Training LLMs can generate a massive amount of data, making communication between devices and the central server resource-intensive. By training models on edge devices, federated learning reduces the need for data transmission over the network, saving bandwidth and lowering communication costs.
- *Distributed computing power*: By promoting decentralised ML, federated learning enables local nodes to participate in the training process. This can improve responsiveness, reduce latency, and enhance autonomy. As a result, leveraging the computing power of multiple devices or servers enables faster training of LLMs by parallelising the training process. Such pooled computational and storage resources federated across a sufficiently large set of participating nodes can even compete with the infrastructural computing resources of the Tech Giants.

To avoid a potential bottleneck and a single point of failure, the community has also proposed so-called *gossip learning* [11], [12] inspired by the gossiping behaviour observed in social networks, which involves the exchange of information and model updates among participating nodes, rather than with one central node. Gossip algorithms are distributed protocols used for information dissemination and aggregation in decentralised systems [13]. In a gossip algorithm, information spreads through the network by means of local peer-to-peer interactions. The algorithm operates in rounds or iterations, and in each round, a node (or a subset of nodes) selects one or more neighbours to exchange information with. Over time, the information spreads across the network as nodes continue to interact and share information with their neighbours. In addition to the default benefits of federated learning, gossip-based extensions provide the following benefits:

- *Scalability*: Gossip learning can scale well with large networks, as each node only needs to communicate with a small number of other nodes at each iteration.
- *Fault tolerance*: Gossip learning is resilient to node failures or network partitions. Even if some nodes become unavailable, the information can still spread through the network via other nodes.
- *Adaptability to dynamic environments*: Gossip learning can adapt to changes in the network, such as nodes joining or leaving dynamically, allowing for continuous learning in evolving environments.

Gossip algorithms provide a decentralised and scalable

[2] https://flower.dev/

[3] https://www.tensorflow.org/federated/

[4] https://github.com/securefederatedai/openfl/

approach for communication in federated learning, allowing devices to collectively learn from each other while preserving data privacy. They can also handle communication failures, device churn, and heterogeneity in device capabilities. Various gossip algorithms, such as random pairwise gossip, ring-based gossip, or hierarchical gossip [14], can be employed depending on the specific requirements and characteristics of the federated learning scenario. The main steps of a federated learning setup, enhanced with gossip algorithms, are the following (depicted in Fig. 2):

1) *Initialisation*: Each participating device initialises its local model with an initial set of parameters.
2) *Local training*: Each device trains its local model using own data, following a predefined training process. This can involve multiple training iterations (*i.e.*, epochs).
3) *Communication and model exchange*: A subset of devices is selected to participate in the communication process. The selection can be random or based on certain criteria, such as device proximity or resource availability. During the communication round, selected devices exchange information, which can include sharing model parameters, gradients, or other relevant updates.
4) *Update aggregation*: Next, each device updates its local model by aggregating the received information from other devices. The aggregation process can vary and may include techniques like averaging, weighted averaging, or more sophisticated aggregation strategies [10].
5) *Repeat*: Steps 2-4 are repeated for multiple communication rounds or until convergence criteria are met. The goal is to iteratively refine the local models and improve the global model without sharing raw data.

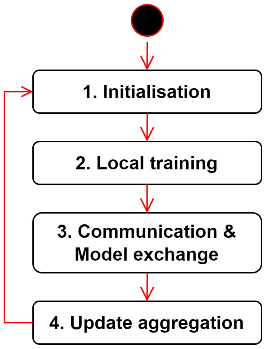

Fig. 2. Federated learning workflow based on crowd-sourced training data.

Foundation for the next layer: Federated learning provides software frameworks for training a global model using private datasets from distributed interconnected devices.

D. Crowd-sourcing

Federated learning makes it technically possible for distributed clients to participate in a collaborative model training process. At the same time, it enables them to safely contribute their local, potentially sensitive datasets for training. In the context of Generative AI, these could be, for example, some technical documentation within an internal corporate network or photo images stored on a personal smartphone. The involvement of human users or organisations in a federated setup is strongly related to the concept of *crowd-sourcing* [15], which can be used for training LLMs by harnessing the collective efforts and knowledge of a diverse group of contributors.

In a broad sense, crowd-sourcing is a process of obtaining services, ideas, or content by soliciting contributions from a large group of people. It involves breaking down a global task into small, discrete parts and distributing those parts among participants, who then upon completion of their local parts, contribute the outputs to achieve the global task. In the context of LLM training, crowd-sourcing would assume the participation of a large and diverse group of online users contributing with their collected private datasets to train a global AI model.

Incentive mechanisms play a crucial role in crowd-sourced federated learning to encourage participation and cooperation among the participating users and organisations [16]. These mechanisms aim to align the interests of the participants with the overall objectives of the federated learning process. Some common incentive mechanisms applicable in this context can be based on monetary, reputation or resource rewards given to the crowd-sourcing participants. Admittedly, designing effective incentive mechanisms is a challenging task, as it requires balancing the objectives of individual participants with the global goals of the Generative AI system, at the same time ensuring fairness, privacy, and end-to-end security.

Another important aspect of crowd-sourcing, provided a sufficiently large number of participating actors, is the *diversity* – *i.e.*, crowd-sourcing allows for the inclusion of diverse perspectives and linguistic variations in the training data, enhancing the language model's ability to handle different languages, cultural contexts, competences, beliefs, etc. By leveraging the collective intelligence and efforts of a diverse group of contributors, crowd-sourcing enables the creation of more robust and inclusive models. Taken together, the sufficiently large number of participants and their diversity in a federated crowd-sourcing setup will enable the so-called *wisdom of the crowd*. The wisdom of the crowd is a concept that suggests that a group of individuals collectively can make better decisions or provide more accurate answers than an individual expert [17]. It is based on the idea that aggregating diverse opinions and knowledge from a large group leads to a more reliable and accurate outcome. The wisdom of the crowd relies on the inclusion of diverse independent viewpoints, opinions, and knowledge. This way, when individuals with different backgrounds and perspectives contribute their insights, it reduces biases and brings a wider range of information to the decision-making process. It is assumed that errors or biases present in individual opinions (*i.e.*, the minority) are cancelled out or outweighed by the collective judgement of the majority. The wisdom of the crowd is a crucial milestone in collective training of unbiased and fair Generative AI models.

Foundation for the next layer: Crowd-sourcing allows using private datasets for training going beyond the publicly available data on the Internet, minimises bias and achieves fair model training results via the wisdom of the crowd.

E. Hierarchical Moderation

A critical aspect of Generative AI is its moderation. AI moderation refers to the process of regulating and controlling the behaviour of AI systems to ensure they operate in a responsible and ethical manner [18]. In an ideal scenario, efficiently implementing crowd-sourcing and achieving the described wisdom of the crowd will assume inherent self-moderation, where the diversity and the large number of participants will ensure that biases and errors of individual training inputs are balanced out by the strengths of others. This can be compared to the majority and plurality rules in democracy – the principles of taking most popular group decisions, where all expressed opinions are treated fairly by giving each an equal weight.

Humanity, however, knows many examples when the majority was wrong. Therefore, a more realistic scenario is the introduction of an additional moderation layer, which will rely on the democratic power separation principles, such that no single authority has the power to evaluate the accuracy of the data or the model (as it happens now with the mainstream Generative AI tools). *Hierarchical moderation* (or hierarchical governance) is a model of content moderation and decision-making that is structured in a hierarchical manner to ensure the quality and reliability of its articles [18]. A notable reference in this context is Wikipedia, which relies on a distributed hierarchy of community-nominated and elected editors to ensure the correctness and fairness of user-generated content. Implementing a similar automated moderation system for crowd-sourced LLMs would rely on training advanced algorithms to detect and remove harmful or offensive content. This could also involve creating separate models that are trained on inappropriate content, allowing to identify similar content and flag it for review. It is important to note that the moderation framework should respect the democratic principles and operate in a collaborative and consensus-driven manner. The community's input and involvement are key to maintaining the integrity and quality of the contents. The hierarchical structure, policies, and roles need to be designed to provide a framework for decision-making and to address the described content moderation and quality control issues.

Noteworthy, maintaining oversight in community-driven AI should not mean reinforcing entrenched biases or imposing a single view of the truth. Rather, it should nurture an environment that enables AI to continually learn, unlearn, and relearn in harmony with the evolving human insights. Essentially, the moderation process should mirror the fluid nature of understanding – a perpetual journey that encourages humility, values diverse perspectives, and nourishes a communal spirit. The focus should be on shared enlightenment and empathy, fostering an AI that augments common human experiences rather than homogenising or belittling them. This oversight includes intensive testing of updates via a validation set, real-time A/B testing, fairness, and bias evaluations, as well as adversarial testing. These tests should be performed by a diverse group of participants using their local validation sets. The results from these tests need to be collated to create a global metric that informs whether new knowledge introduced into the community-driven AI requires modification or complete exclusion.

Foundation for the next layer: Moderation enables producing fair and accurate language models with minimum bias or misinformation. This way, the models can be further safely used in various applications.

F. Applications

Finally, with the rest of the elements in place, it will be eventually possible to build the Generative AI software tools based on the collaboratively-trained models. The scope of such tools is not expected to differ from the currently being in use. While new applications and use cases continue to emerge on a daily basis, LLMs are already playing a key role in the following scenarios: various chat bots, virtual assistants and recommendation systems, content comprehension, generation and moderation, sentiment analysis and opinion mining, to name a few. It is expected that the unbiased and transparent nature of the community-driven Generative AI tools will create fair competition with the proprietary commercial tools, thus also facilitating the increased quality of the available products.

Coming back to the motivating example, the proposed community-driven Generative AI could provide an effective solution to build a more accurate and intelligent assisted coding functionality. Leveraging both crowd-sourcing and federated learning, organisations and individuals can contribute their unique knowledge to the AI model, including proprietary coding methodologies that may be less common in public and open-source code. This non-public data could come from private repositories, issue trackers and enterprise source code management systems. Federated learning ensures privacy of data by only sharing the updated model parameters, thereby preserving the privacy of proprietary and sensitive information. Simultaneously, the model becomes enriched with a diverse range of programming knowledge, extending beyond what public repositories can offer.

The practical implications of such an approach in the field of programming could be transformative. Potential benefits include more enhanced problem-solving capabilities, a richer understanding of proprietary coding practices and improved knowledge of lesser used programming languages. This community-driven model could democratise AI within the programming sphere, allowing even smaller contributors to play a role in AI development. Careful considerations of intellectual propoerty rights and the fostering of a robust community to drive this process forward are necessary for the successful implementation of this model. As such, the proposed community-driven Generative AI provides a promising alternative that not only addresses current limitations but also promotes fair competition with proprietary commercial tools.

III. Assumptions and Further Considerations

In this section, we discuss our concerns of non-technical nature, which should also be taken into account, as well research considerations going beyond the scope of this paper.

A. Role of the Community

Combined, the described technologies represent a powerful toolkit for building a community-driven Generative AI ecosystem, which can challenge the establishing plutocracy of big corporations. We have already seen examples of similar large-scale collaborative projects in the past. The most notable example is NASA's SETI@home project [19], which connected more than 5 million users contributing their private computing resourcing. A more recent example using federated learning is the MELLODY project,[5] which connected several medical research institutions into a federated learning network used for drug discovery in the pharmaceutical industry.

Also, the proposed vision shares many similarities with various open-source software foundations, such as Linux, Apache, and Eclipse. All these organisations are community-driven and follow the principles of transparent and open communication and code distribution. Therefore, an important assumption of this proposed vision is the active involvement of the community in establishing and further developing such an open ecosystem for Generative AI. As we already argued, even in the presence of crowd-sourcing and decentralised gossip learning, there is still a need for aggregating the model, developing the training algorithms, as well as the moderation – all these activities cannot (and should not) be fully decentralised.

B. Intellectual Property Rights

Intellectual property rights (IPR) is another important consideration in the context of envisioned architecture, as it involves collaboration and sharing of information among multiple parties, followed by generation of new creative content and its eventual consumption by end users. While there is still many open questions, the key considerations related to IPR can be summarised as follows:

- *Data and model ownership*: In a federated learning setup, the participants typically retain ownership of their data. This means that the data used for training the models remains under the control and ownership of the participants. At the same time, the resulting global model itself may be subject to IPR. The ownership of the model can vary depending on the agreements and arrangements between the parties involved. It is the responsibility of the community to establish clear guidelines and agreements regarding the ownership and use of the trained models.
- *Copyright*: Copyright law protects original creative works, such as text, images, music, or videos, from unauthorised copying, distribution, or use. In the case of Generative AI, questions arise regarding the ownership of AI-generated content. Typically, the copyright ownership is attributed to the human creator or the owner of the AI system, as they provide the input and training data for the AI model. In the proposed architecture, however, there is no such single actor. Therefore, it is again up to the community to decide and agree on the applicable copyright ownership policies.
- *Derivative works*: Generative AI models can be used to create derivative works based on existing copyrighted material (*i.e.*, private data crowd-sourced for federated learning). The legal implications associated with the creation and use of such derivative works need to be further explored, as they may require permission or licensing from the original copyright holder.

The application of IP laws to generative AI raises complex and evolving legal questions. Different jurisdictions may have different interpretations and regulations regarding ownership, copyright, and patentability of AI-generated works. As the whole field continues to advance, legal frameworks are also evolving to address the emerging challenges and opportunities.

C. Wisdom of the Crowd vs Epistemological Relativism

The responses generated by Generative AI, and LLMs in particular, are based on statistical patterns and associations learned from vast amounts of training data. While the models themselves do not have the ability to directly assess the majority opinion or conduct a voting process about what is true and what is false, the crowd-sourcing method used to collect the training data may rely on the previously described wisdom of the crowd principle. By its nature, it assumes that the sufficient number and diversity of contributors will ensure that individual biases and possible errors will be levelled out by the rest of the contributors. This can be seen as a strong assumption, especially from the epistemological relativism point of view. *Epistemological relativism* is a philosophical position, according to which knowledge and truth are not absolute, universal, or objective, but are instead relative to specific individuals, cultures, societies, or historical contexts [20]. In other words, different perspectives, beliefs, and interpretations can be equally valid and legitimate, depending on the context in which they arise. Epistemological relativism challenges the notion of universal truths and emphasises the role of individual perspectives and cultural contexts in shaping knowledge and truth. It focuses on the diversity of subjective interpretations. The wisdom of the crowd, on the other hand, suggests that aggregating diverse perspectives and independent judgements can lead to a single, more accurate outcome.

These two conflicting viewpoints again highlight the need for a thoroughly designed community-driven moderation framework in the proposed architecture. The outputs of LLMs are influenced by the distribution of the training data, inevitably containing certain biases and limitations. Even in the presence of a moderation framework, when using LLMs in applications involving decision-making or opinion representation, it is crucial to consider the limitations and potential biases in the training data and to supplement the outputs with appropriate human judgement and critical evaluation.

[5] https://www.melloddy.eu/

D. Bio-inspired Decision Making and Moderation

One possible way of enhancing the conventional AI moderation is to enhance it existing approaches from other relevant scenarios. Multiple crowd-sourcing contributors participating in a federated learning setup can be seen as individual agents providing their individual information into the global shared pool. Research approaches in the direction of multi-agent systems [21] developed many decentralised decision-making mechanisms inspired from malty-party auctions, arbitration, *etc.* Swarm intelligence [22] learns how advanced intelligence emerges from a swarm of low-intelligent individuals. Furthermore, as a step towards Artificial General Intelligence, community-driven Generative AI could also benefit from even more advanced mechanisms to address the challenge of decentralised decision making and moderation, and look into bio-inspired approaches and, more specifically, into how such processes are organised within a human brain.

Although it may sound counter-intuitive, a human brain is a highly decentralised system. A commonly accepted modern theory explains a brain's decision making based on a *global neuronal workspace* [23]. When individual inputs from the body arrive, multiple local processors within the brain autonomously perform continuous analysis of different parts of the input, also in combination with the local knowledge available to them. The results from these local processors are then 'broadcast' to a global neuronal workspace, from where other local processors can receive them. These other local processors may (or may not) find certain shared results interesting, pick them up for further evaluation and eventually place them back into the workspace. This way, certain results become more 'popular' than others, resulting in more and more local processors noticing and picking them up, and eventually becoming the final decision adopted by the whole brain. This process is to a great extent again similar to the majority and plurality principles observed in modern democracy. Applying this bio-inspired global workspace theory to the community-driven Generative AI for decentralised decision making and moderation might be an interesting and promising direction, which is however still at a very early stage of explorations [24]. Many fundamental challenges remain unaddressed, such as how to effectively broadcast local results, how to attract attention of relevant agents, how to evaluate and define the winning majority of the results, *etc.*

E. Federated Machine Unlearning

Within the sphere of community-driven AI, the elimination of unwelcome content already present in the trained models is of paramount importance. Sources of such undesirable content could be manifold. For example, adversarial data, when used in the training of generative AI models, can infuse inaccurate information into the resulting model. Likewise, adversarial attacks might leak confidential data kept within the model or users might occasionally produce data, such as inadvertent search queries leading to incorrect recommendations. Furthermore, community-driven AI systems, when trained on public datasets, often unintentionally inherit deep-seated racial and cultural biases. All such undesirable content contributes to encouraging biases, spreading harm, and eroding human dignity, and therefore should be wiped out from the models.

To this end, another important aspect of community-driven AI moderation is *unlearning* [25], [26] – *i.e.*, excluding some training results after they have already been included in the model. Federated machine unlearning represents the collective process of detecting sensitive and inaccurate predictions in a community-driven AI model and collaboratively unlearning the information housed within the model. This collaboration might commence with an individual's proposal to unlearn a category, a sample, a task, user-contributed data, or a data stream created by the AI. Once a proposal is tabled, the community should embark on an open and transparent process of consensus-building regarding what to unlearn and the specific methodology to follow. Every stage of the consensus-building journey for each proposal should be documented and retrievable. Upon reaching a consensus, new datasets for training should be constructed to revise what was previously learned. As an example, Wu *et al.* [27] delve into federated machine unlearning by reversing the stochastic gradient descent process for training and implementing elastic weight consolidation. After the fine-tuning/training phase, it is crucial to confirm and quantify the level of unlearning [28] achieved within the updated AI model. The data generated for 'forgetting' should challenge the model to reveal sensitive and incorrect information intended to be forgotten. Metrics should assess the degree of forgetting realised by the model in a recurring unlearning process. Federated machine unlearning is an innovative and emerging field where building consensus and unlearning present considerable challenges.

IV. Conclusion

In this paper, we aimed to challenge the establishing monopoly of the Big Tech on the Generative AI market by proposing a conceptual architecture for community-driven Generative AI. The envisioned architecture consists of several technological building blocks, among which we consider crowd-sourcing and federated learning to be the main enablers. By soliciting training data from a wide range of contributing parties, crowd-sourcing can capture a range of opinions, insights, and expertise that might otherwise be missed, thus providing a more comprehensive and unbiased view on a topic than any individual or organisation can offer. By drawing on a wider pool of perspectives and experiences this way, crowd-sourcing can help achieving the so-called wisdom of the crowd, where the collective intelligence arises from the aggregation of individual opinions, perspectives, and experiences, which can cancel out errors and biases and lead to more accurate and robust outcomes. At the same time, federated learning will ensure that data will remain private, since it assumes that instead of sending data samples to the central server, each client performs local training on its own data, and only exposes the updated model parameters which are then aggregated and shared back to the participating clients.

We have also considered several assumptions and potential research directions which still need further investigations. These include the important role of the community that will drive the whole development process, the IPR implications associated with the AI-generated content, the general epistemological concerns of the knowledge used to train the AI models, and finally possible ways of enhancing the AI moderation system by applying bio-inspired decision making and federated machine unlearning. Addressing all these open questions requires input from multiple stakeholders, including technologists, researchers, content creators, and ordinary users, as well as policy makers and civil society organisations.

Acknowledgment

This work has received funding from the European Union's HEU and H2020 research and innovation programmes under grant agreements No. 101095634 (ENTRUST) and No. 101020416 (ERATOSTHENES), and from the Research Council of Norway's BIA-IPN programme under grant agreement No. 309700 (FLEET).

References

[1] V. Talla, M. Hessar, B. Kellogg, A. Najafi, J. R. Smith, and S. Gollakota, "LoRa Backscatter: Enabling The Vision of Ubiquitous Connectivity," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 1, no. 3, pp. 1–24, 2017, doi: https://doi.org/10.1145/3130970.

[2] M. R. Ebling, "Pervasive Computing and the Internet of Things," *IEEE Pervasive Computing*, vol. 15, no. 1, pp. 2–4, 2016, doi: https://doi.org/10.1109/MPRV.2016.7.

[3] R. Dautov and S. Distefano, "Three-level hierarchical data fusion through the IoT, edge, and cloud computing," in *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*. ACM New York, NY, USA, 2017, pp. 1–5, doi: https://doi.org/10.1145/3109761.3158388.

[4] R. Dautov, S. Distefano, D. Bruneo, F. Longo, G. Merlino, and A. Puliafito, "Pushing intelligence to the edge with a stream processing architecture," in *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, 2017, pp. 792–799, doi: https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2017.121.

[5] X. Wang, Y. Han, V. C. Leung, D. Niyato, X. Yan, and X. Chen, *Edge AI: Convergence of edge computing and artificial intelligence*. Springer, 2020, doi: https://doi.org/10.1007/978-981-15-6186-3.

[6] E. J. Husom, R. Dautov, A. Nedisan Videsjorden, F. Gonidis, S. Papatzelos, and N. Malamas, "Machine Learning for Fatigue Detection using Fitbit Fitness Trackers," in *Proceedings of the 10th International Conference on Sport Sciences Research and Technology Support - icSPORTS*, INSTICC. SciTePress, 2022, pp. 41–52, doi: https://doi.org/10.5220/0011527500003321.

[7] R. Dautov, E. J. Husom, F. Gonidis, S. Papatzelos, and N. Malamas, "Bridging the Gap Between Java and Python in Mobile Software Development to Enable MLOps," in *2022 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2022, pp. 363–368, doi: https://doi.org/10.1109/WiMob55322.2022.9941679.

[8] R. Dautov, S. Distefano, G. Merlino, D. Bruneo, F. Longo, and A. Puliafito, "Towards a Global Intelligent Surveillance System," in *Proceedings of the 11th International Conference on Distributed Smart Cameras*. ACM New York, NY, USA, 2017, pp. 119–124, doi: https://doi.org/10.1145/3131885.3131918.

[9] R. Dautov, S. Distefano, D. Bruneo, F. Longo, G. Merlino, A. Puliafito, and R. Buyya, "Metropolitan intelligent surveillance systems for urban areas by harnessing IoT and edge computing paradigms," *Software: Practice and experience*, vol. 48, no. 8, pp. 1475–1492, 2018, doi: https://doi.org/10.1002/spe.2586.

[10] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," in *29th Conference on Neural Information Processing Systems (NIPS2016)*, 2016, pp. 1–5, doi: https://doi.org/10.48550/arXiv.1610.05492.

[11] I. Hegedűs, G. Danner, and M. Jelasity, "Gossip Learning as a Decentralized Alternative to Federated Learning," in *Distributed Applications and Interoperable Systems (DAIS 2019)*, J. Pereira and L. Ricci, Eds. Springer, 2019, pp. 74–90, doi: https://doi.org/10.1007/978-3-030-22496-7_5.

[12] G. Li, Y. Hu, M. Zhang, L. Li, T. Chang, and Q. Yin, "FedGosp: A Novel Framework of Gossip Federated Learning for Data Heterogeneity," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2022, pp. 840–845, doi: https://doi.org/10.1109/SMC53654.2022.9945192.

[13] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Gossip algorithms: Design, analysis and applications," in *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, vol. 3. IEEE, 2005, pp. 1653–1664, doi: https://doi.org/10.1109/INFCOM.2005.1498447.

[14] D. Shah, "Gossip algorithms," *Foundations and Trends® in Networking*, vol. 3, no. 1, pp. 1–125, 2009, doi: https://dx.doi.org/10.1561/1300000014.

[15] H. O. Ikediego, M. Ilkan, A. M. Abubakar, and F. V. Bekun, "Crowdsourcing (who, why and what)," *International Journal of Crowd Science*, vol. 2, no. 1, pp. 27–41, 2018, doi: https://doi.org/10.1108/IJCS-07-2017-0005.

[16] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *2012 Proceedings IEEE INFOCOM*. IEEE, 2012, pp. 1701–1709, doi: https://doi.org/10.1109/INFCOM.2012.6195541.

[17] I. Kremer, Y. Mansour, and M. Perry, "Implementing the "wisdom of the crowd"," *Journal of Political Economy*, vol. 122, no. 5, pp. 988–1012, 2014, doi: https://doi.org/10.1086/676597.

[18] T. Gillespie, "Content moderation, ai, and the question of scale," *Big Data & Society*, vol. 7, no. 2, p. 2053951720943234, 2020, doi: https://doi.org/10.1177/2053951720943234.

[19] D. P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer, "SETI@home: an experiment in public-resource computing," *Commun. ACM*, vol. 45, no. 11, pp. 56–61, 2002, doi: https://doi.org/10.1145/581571.581573.

[20] S. Luper, "Epistemic relativism," *Philosophical Issues*, vol. 14, pp. 271–295, 2004, doi: http://dx.doi.org/10.1111/j.1533-6077.2004.00031.x.

[21] A. Dorri, S. S. Kanhere, and R. Jurdak, "Multi-Agent Systems: A Survey," *IEEE Access*, vol. 6, pp. 28 573–28 593, 2018, doi: https://doi.org/10.1109/ACCESS.2018.2831228.

[22] A. Chakraborty and A. K. Kar, "Swarm Intelligence: A Review of Algorithms," in *Nature-inspired computing and optimization: Theory and applications*, S. Patnaik, X.-S. Yang, and K. Nakamatsu, Eds. Springer, 2017, pp. 475–494, doi: https://doi.org/10.1007/978-3-319-50920-4_19.

[23] G. A. Mashour, P. Roelfsema, J.-P. Changeux, and S. Dehaene, "Conscious Processing and the Global Neuronal Workspace Hypothesis," *Neuron*, vol. 105, no. 5, pp. 776–798, 2020, doi: https://doi.org/10.1016/j.neuron.2020.01.026.

[24] R. VanRullen and R. Kanai, "Deep learning and the global workspace theory," *Trends in Neurosciences*, vol. 44, no. 9, pp. 692–704, 2021, doi: https://doi.org/10.1016/j.tins.2021.04.005.

[25] H. Zhang, T. Nakamura, T. Isohara, and K. Sakurai, "A Review on Machine Unlearning," *SN Computer Science*, vol. 4, no. 4, p. 337, 2023, doi: https://doi.org/10.1007/s42979-023-01767-4.

[26] H. Xu, T. Zhu, L. Zhang, W. Zhou, and P. S. Yu, "Machine unlearning: A survey," *ACM Computing Surveys*, 2023, doi: https://doi.org/10.1145/3603620.

[27] L. Wu, S. Guo, J. Wang, Z. Hong, J. Zhang, and Y. Ding, "Federated Unlearning: Guarantee the Right of Clients to Forget," *IEEE Network*, vol. 36, no. 5, pp. 129–135, 2022, doi: https://doi.org/10.1109/MNET.001.2200198.

[28] X. Gao, X. Ma, J. Wang, Y. Sun, B. Li, S. Ji, P. Cheng, and J. Chen, "VeriFi: Towards Verifiable Federated Unlearning," *Computing Research Repository*, 2022, doi: https://doi.org/10.48550/arXiv.2205.12709.