

# The BitRobot Network

## Whitepaper

Michael Cho, Jonathan Victor, Juan Benet

Mar 10, 2025

# Table of Contents

## 1. Abstract

---

## 2. Introduction

---

## 3. Problem Definition

- 3.1 Data Challenges
  - 3.2 Resource Challenges
- 

## 4. The BitRobot Network: Design and Protocol

- 4.1 Overview of the BitRobot Network
  - 4.2 Key Components of the Network
    - Verifiable Robotic Work (VRW)
    - Embodied Node Token (ENT)
    - Subnets:
      - Subnet Owners
      - Subnet Validator
      - Subnet Contributors
  - 4.3 Life cycles of Subnets and Robots
  - 4.4 IP Assignment and Ownership
  - 4.5 Governance
    - BitRobot Foundation
    - BitRobot Senate
    - Gandalf AI
    - Network Participants
- 

## 5. BitRobot Economy

- 5.1 Economic Flows to Subnets
  - 5.2 Economic Consumption
  - 5.3 Economic Loops
- 

## 6. Future Work

---

## 7. References

# 1. Abstract

The BitRobot Network whitepaper proposes a subnet-based architecture designed for distributed robotic work and collaboration. Key components include Verifiable Robotic Work (VRW) for robotic task definition and verification, Equipment Node Tokens (ENT) for equipment ownership and network access, and Subnets as the operational layer for task execution. Subnets are managed by Subnet Owners, who oversee their operation; Subnet Validators, who verify task completion; and Subnet Contributors, who supply equipment and resources. The economic model aims to align incentives across participants and facilitates seamless coordination. This framework enables the aggregation of critical resources, with the aim of driving innovation and progress in the field of Embodied AI.

## 2. Introduction

The arms race for Artificial General Intelligence (AGI) is in full motion - with companies and governments racing to acquire the data, compute and talent required to create the most performant frontier models. Given the pace of innovation for AI in the digital context (e.g. for text, images, and videos), there are high hopes that a similar step change in growth is around the corner for the world of robotics. This optimism is bolstered by momentum across several critical trends:

**Transformers scaling across modalities:** The transformer architecture [1] has proven versatile in its ability to scale both in quantity, with large amounts of data and compute, and scope, across a range of modalities (e.g. text, images, and videos) - leading to a single unified model that works across multimodal domains (e.g. VLMs [2]). Robotics data, when thought of as transformer tokens, can fit within this same architecture, and therefore should receive scaling benefits when combined with other modalities (e.g. VLAs [3]).

**Learning from demonstrations:** Human teleoperated demonstrations (e.g. techniques like behavioral cloning, imitation learning, offline reinforcement learning, etc) have begun to show real promise. Tasks that the field has previously struggled with [4, 18] (e.g. clothes folding) are starting to see positive momentum - and if scaled may help solve for generalization of challenging domains (e.g. dexterous manipulation).

**Advancements in simulation:** Simulations have evolved substantially in certain domains (e.g. locomotion), where policies trained in simulations can often transfer zero-shot to real-world embodiments such as quadrupeds [5] and drones [6]. The rise of generative AI, the increasing number of realistic world models [7, 8, 9] and the proliferation of physics engine toolings [10, 11, 12] all point to the potential for using simulations to accelerate progress by building a more robust foundation for models. .

**Learning from videos:** Web-scale videos also point towards a rich, largely untapped, reservoir of training data that can help advance Embodied AI. In particular, human videos, especially those filmed with ego-centric views, show promise for humanoids and bimanual robotic arms [13, 14]. Meanwhile, progress is also being made where human videos may be closing the gap with labor-intensive teleoperated robotic data [15].

**Learning from cross embodiment:** Increasing research evidence indicates that there may be significant positive learning transfer when Embodied AI models are trained from a diverse range of robotic embodiments. This suggests that the absolute quantity of data required may be substantially reduced and that the resulting models may be more generalizable if the training datasets come from a wide array of robot types [16, 17].

**Advancements in robotic hardware:** The maturity of adjacent industries (e.g. smartphones, EVs, semiconductor chips) translates to strong supply chains and cheap components for inputs to robotic systems (e.g. actuators, cameras, GPS, IoT sensors). For certain types of robots (e.g. consumer-grade sidewalk robots, robotic arms, quadrupeds, etc), this means affordable hardware that can be distributed at scale at consumer price points. In parallel, there are increased innovations on new hardware designs, ranging from tendon-driven mechanics, pneumatics and hydraulic-based actuations and various new touch/tactile sensors.

## 3. Problem Definition

Despite the momentum described above, Embodied AI still faces many real-world constraints that make scaling of physical intelligence a significant bottleneck in the pursuit of true AGI.

### 3.1 Data Challenges

Below we describe the challenges facing the three main types of robotic data (synthetic data, real-world video data, and teleoperation data) needed for improving Embodied AI.

**Sim-to-real gap:** Despite its success in certain subfields, synthetic data has had challenges transferring zero-shot in real world settings for many other robotic tasks (e.g. in-the-wild robotic navigation, robotic manipulation with deformable objects). While simulations can drastically speed up the creation of training scenarios, the challenge is in focusing the search space to the long tail of edge cases that are actually relevant in the real world.

**Embodiment gap:** Real-world video data, while a promising accelerant to Embodied AI research, suffer from the embodiment gap. Video data lacks precise information about physical presence and contextual awareness, making it difficult to build an AI that can replicate the exact responses a human might have. For example, thousands of video hours of Lionel Messi playing football can build a basic foundation, but lacks information about his specific body, data processing, and sensory inputs - making it challenging to exactly replicate his actual decision making and execution on the field.

**Unconscious-to-conscious gap:** Unlike other modalities (e.g. text) which are available at scale on the public internet, teleoperation data, collected in real world settings capturing human action inputs as a response to external events, requires costly coordination and investment of both physical hardware and human man-hours to produce. As a result, while a

promising data source for Embodied AI models, very few large-scale teleoperation datasets naturally exist [19].

Critically, while many robotics researchers agree that these are the primary data bottlenecks to solving Embodied AI, there is a lack of consensus about which challenge (or combination of challenges) are the most pressing to solve first.

### 3.2 Resource Challenges

Furthermore, while robotics data remains a significant bottleneck, real-world evaluations and resource asymmetries may prove to be an even larger point of friction to the pace of Embodied AI innovations.

**Lack of hardware for real-world evaluations:** For Embodied AI, the only way to evaluate a new model is to observe how the model fares in real world settings. Witnessing failure cases (e.g. a self-driving car driving into a lamp post) is critical in understanding the limits to what a model can do - and as the model improves, the rate of failure is only observable with large-scale fleets of robots. As a result, the ability for researchers to evaluate their own models is rate limited, without access to large amounts of robots to which to deploy their models.

**Resource distribution asymmetry:** While a few large corporate labs are able to make substantial progress (given they have access to dataset flywheels, evaluation platforms, and compute, storage and bandwidth) - large pools of researchers (in academia or early-stage startups) and hardware providers are unable to iterate at the same pace. This points to a worrying trend of significant concentration of scientific progress in just a handful of entities.

Below is an illustration of the asymmetry of access to various resources amongst the various entities.

	Tesla	Deepmind	Large Labs	Startups	Academia
Research Talent	✓	✓	✓	✓	✓
Large Robot Fleet	✓ (Cars)				
Compute & Storage	✓	✓	✓	✓	
Teleoperation Dataset	✓ (Drivers)				
Synthetic Dataset	✓	✓	✓	✓	
Video Dataset		✓ (YouTube)			

Fig 1: Rough mapping of resource access for key players in the Embodied AI landscape.

# 4. The BitRobot Network: Design and Protocol

## 4.1 Overview of the BitRobot Network

In response to the challenges listed above, we present the BitRobot Network, a modular network of robotic subnets, each generating valuable resources, to accelerate Embodied AI and robotics innovation.

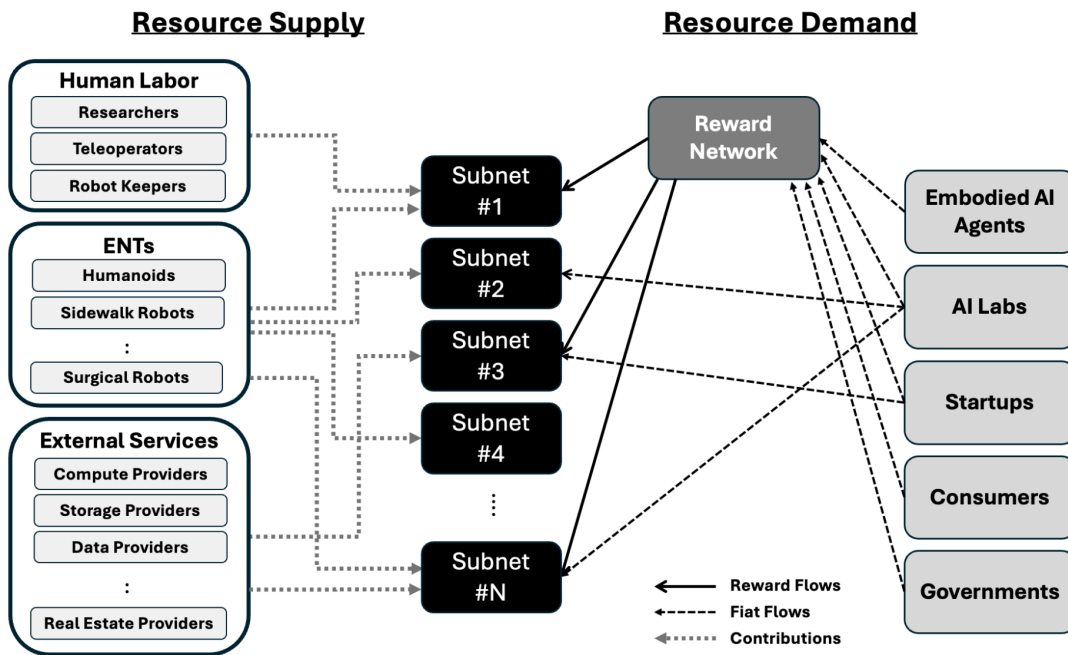


Fig 2: Contribution and Compensation Flows for the BitRobot Network. Note: Compensation flows from Subnets to their Contributors are Subnet defined and not shown.

In the BitRobot Network, subnets can be formed to solve for resource gaps on the path to Embodied AI. This can range from aggregating hardware (e.g. fleets of robots for massively parallelized evaluation of new AI models) or human labor (e.g. teleoperators to control various robots) to more sophisticated coordination among multiple resources (e.g. researchers testing latest models on real world robot fleets while training on GPUs provided by compute providers). By decentralizing resource aggregation, BitRobot can leapfrog the resources that any individual corporate lab might amass on its own - and democratize access across the full Embodied AI community.

The subnet structure allows for permissionless resource formation and experimentation - while aligning all participants around a shared economy and outcome.

## 4.2 Key Components of the BitRobot Network

The BitRobot Network is built around the following components:

**Verifiable Robotic Work (VRW):** The Verifiable Robotic Work (VRW) is a quantifiable and

verifiable measure of any robotic work that is deemed useful to help advance Embodied AI. This may include work done collecting data in the real world (e.g. the total distance driven by sidewalk robots, the difficulty scores of these sidewalk robot drives, etc) or in simulation (e.g. the number of scenarios generated for robotic arms that are tasked to fold clothes, the quality scores of such generated scenes, etc). VRW may also include the work done in creating Embodied AI models and evaluating them in either real-world settings or simulation environments. Finally, different types of Subnet Contributors may have different VRW even within the same Subnet (e.g. teleoperator based on duration of operating robots, robot keepers based on robot online hours, etc).

**Embodied Node Token (ENT):** The Embodied Node Token (ENT) is a unique identifier within the system, designed to serve as the digital twin of a physical robot. Represented as an NFT, the ENT ensures a unique identity for each robot in the BitRobot Network. Each ENT requires collateral to receive on-chain payments, which can be used to enforce accountability and incentivize compliance within the system. The ENT also functions as the conduit for payments for robot owners.

**Subnets:** A BitRobot Subnet represents clusters of resources creating value for the BitRobot Network. Subnets can range in scope: from hardware acquisition, to creation of novel datasets, to creation of new AI models, to competitions. Subnets can work on a diverse range of tasks (defined by each Subnet's *Verifiable Robotic Work*) and embodiments, resulting in a wide variety of outputs, ranging from real-world datasets, synthetic datasets, to new AI models, physical hardware fleet, etc. Within Subnets there are several roles: Subnet Owners, Subnet Validators, and Subnet contributors - though many participants may play multiple roles and work across multiple subnets. Subnets can earn income via direct payments (e.g. stablecoins for renting out a fleet of robots), by network emissions, or both.

**Subnet Owners:** Subnet Owners represent the creators of a subnet. Subnet owners must define the *VRW* for their subnet, the type of subnet outputs (e.g. dataset, AI models, cryptographic proofs), subnet validation, and the percent of subnet rewards to be distributed to the various participants in the Subnet. Subnet Owners can be individuals, groups of people, or protocols.

**Subnet Validators:** Subnet Validators evaluate the output of Subnet Contributors against the defined *VRW* for the subnet. Internal to the subnet, the scores assigned by Subnet Validators determine how incoming subnet rewards are split among Subnet Contributors. Subnets can have as few or as many validators as the Subnet Owners determine, and Subnet Validators may be Subnet Owners themselves.

**Subnet Contributors:** Subnet Contributors provide the necessary inputs for a subnet to create a useful output. Each subnet may have multiple types of contributors - ranging from researchers, to teleoperators, to robot manufacturers, and many more. Given each Subnet may be focused on a different outcome, it is expected that each subnet will have a different set of contributor types based on the requirements for their tasks.







Subnets	Embodiment	Verifiable Robotic Work	Subnet Contributors					Subnet Output
			Teleoperators	Robot Owners	Robot Keepers	Compute Providers	Researchers	
SN1	 Sidewalk Robots	Human-Teleoperated Urban Navigation	✓	✓	✓			Real-World Dataset
SN2	 Sidewalk Robots	Autonomous Urban Navigation Models		✓	✓	✓	✓	AI Models
SN3	 Humanoids	Human-Teleoperated Cooking	✓	✓	✓			Real-World Dataset
SN4	 Humanoids	Provision of Humanoid Fleet		✓	✓			Physical Hardware
SN5	 Humans	Provision of Human Teleoperators	✓					Human Physical Intelligence
⋮								
SN ?	 Quadruped Robots	Simulation of Large Fleet Movement				✓	✓	Synthetic Dataset

Fig 3: Illustrative map of possible Subnets and their corresponding Contributors and Outputs

## 4.3 Life cycles of Subnets and Contributors

### 4.3.1 Subnets

Subnets are instantiated by their owners, and are required to define several parameters including: reward splits, VRW definitions, and validation conditions. Subnets can be constructed to aggregate resources (e.g. humanoid robot pools for hire), to create novel datasets (e.g. urban navigation data), or to simply evaluate models.

Subnets can also be defined to be private or public - private subnets are not eligible for emissions, while public subnets are eligible. Private subnets are consumers of the frameworks and resources of the BitRobot Network, while public subnets may both be consumers and contributors (see Section 4.4 for more details). Public subnets will receive their share of rewards based on the delegation-weighted voting mechanism at every voting epoch. Funds received by subnets are split to reward Subnet Owners, Validators and Contributors based on each subnet's predefined allocation rules.



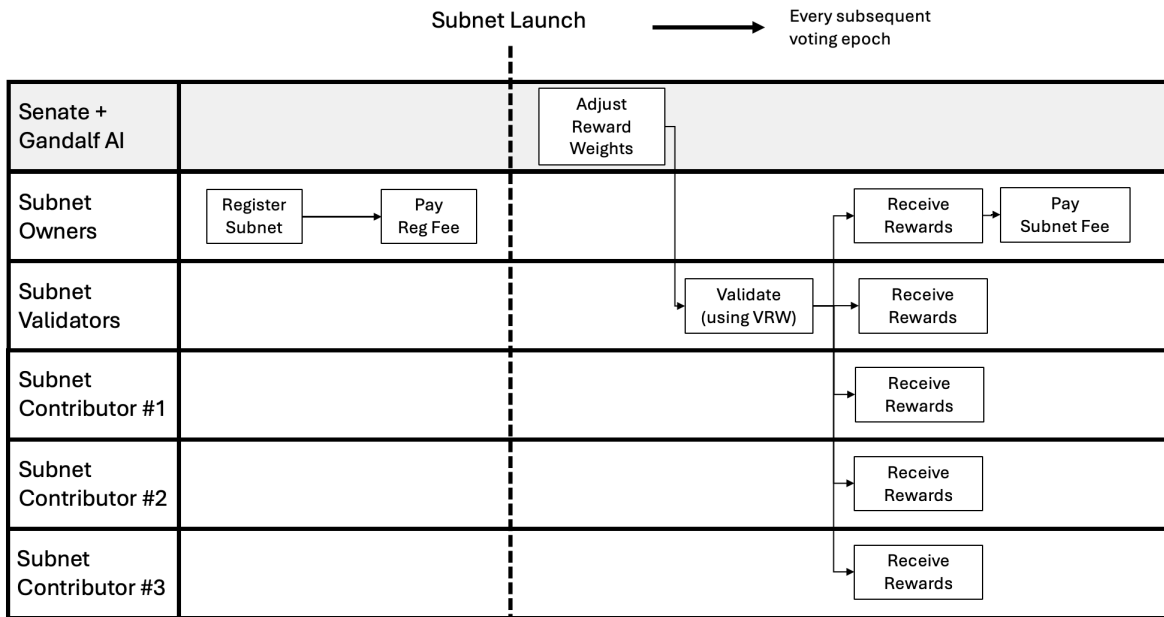


Fig 4: Flowchart showing the major milestones from creation to the ongoing running of a public subnet

#### 4.3.2 ENTs and Contributors

Once subnets are spun up, contributors can join subnets to participate in the defined forms of verifiable work. For human contributors, this can be as simple as taking on tasks available from the various subnets (e.g. teleoperating a robot) using a wallet as an identifier. For robot contributors, they first need to be registered with an ENT NFT. Robot owners (who own ENT NFTs) will be able to participate in multiple relevant subnets while maximizing the utilization rate of each physical robot.

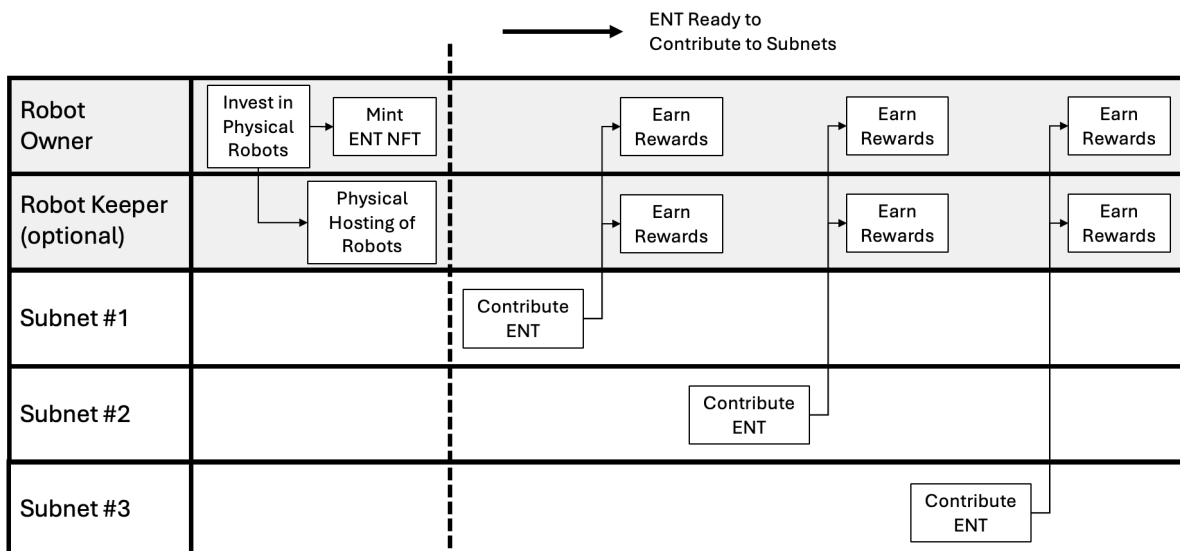


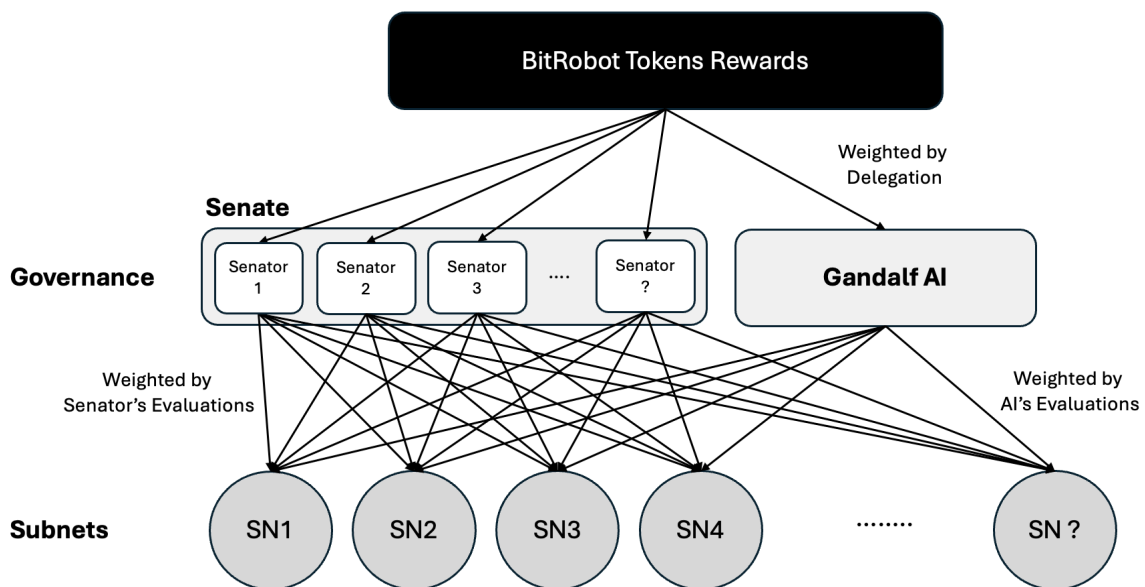
Fig 5: Flowchart showing the major milestones for adding a robotic resource to a subnet

## 4.4 IP Assignment and Ownership

Datasets and Models produced by public subnets (emission-receiving) in the BitRobot Network are by default open-sourced for non-commercial use.

Any datasets or models produced by private subnets (non-emissions receiving) do not fall under this regime.

## 4.5 Governance



## BitRobot Foundation

The BitRobot Foundation's mission is to advance Embodied AI research by facilitating the growth of the BitRobot Network.

To achieve this mission, the BitRobot Foundation will help grow and support infrastructure development work (e.g. tracking of work done by individual ENTs, networking support for subnets working on teleoperation, data pipeline support for subnets working on datasets generation, etc), providing administrative support for the Senate and evangelizing work (e.g. hackathons, grant administrations, etc), and supporting grants to develop the supporting ecosystem.

The BitRobot Foundation will also take on more Embodied AI research and engineering work over time in order to fully maximize the research and economic impacts arising from the commercial use of its IPs (datasets or AI models).

## BitRobot Senate

The BitRobot Senate is composed of nominated representatives of the broader

cryptocurrency and Embodied AI research communities - aimed at ensuring resources flow towards the subnets creating the highest value. The BitRobot Senate governs economic weights associated with network measurements (see Section 5.1 below) which are used to evaluate the value of the outputs generated by the subnets.

At regular intervals (called "Voting Epochs"), the BitRobot Senate will vote on the inclusion of new robotic subnets for emissions. At the same time, the BitRobot Senators will also individually vote on the weights proposals for network distributions assigned to each robotic subnet based on the performance and value contribution from the individual subnet to the overall network.

## Gandalf AI

The BitRobot Foundation will develop and operate "Gandalf AI", an ongoing effort to create a fully open-sourced AI agent model that provides weights proposals for network measurements and distributions for each robotic subnet, serving as a counter balance to the economic influence wielded by the Senate.

## Network Participants

BitRobot Network participants can directly influence the impact of decisions made by Senators and Gandalf AI via delegation of their voting power. In this way, the participants can provide a check on any individual senator by allocating more decision making power with other senators or in an extreme case - entirely allocating decision making power to Gandalf AI.

# 5. BitRobot Economy

The BitRobot Economy is composed of a diverse set of participants from all stages of the value creation pipeline (early research, fundamental development, productionisation). The structure of the BitRobot economy aims to support all of these participants by balancing the flows of the economy across the different types of contribution to the value chain.

## 5.1 Economic Flows to Subnets

The BitRobot Senate and Gandalf AI will use a Measurement-Evaluation-Rewards (MER) loop to determine how emissions should be allocated between subnets. As described in the next section, BitRobot Network participants can increase (or decrease) the relative power of the BitRobot Senate or Gandalf AI's votes via delegation.

**Measurements:** Defined via protocol governance, these are the metrics (and its relevant calculation) that subnets will be evaluated against. For example, a measurement may apply to a specific subnet (e.g. "Qualified Miles" for Subnet 1) or a more generic measurement that could be relevant to multiple subnets (e.g. number of rented humanoid robots). Members of

the Senate and Gandalf AI will individually assign weights to each measurement, indicating their opinion of the relative importance of each metric against the rest.

**Evaluation:** Members of the Senate and Gandalf AI will individually evaluate subnets against each of the relevant metrics to create a distribution list per Member based on their own assessment.

**Reward:** Each Senator and Gandalf AI's distribution lists are weighted by their delegations, summed, and normalized to produce the distribution for that epoch of emissions to each subnet.

Specifically,

$D_j$  : Weighted delegations to Gandalf AI or Senator  $j$

$N_l$  : New emissions in Voting Epoch  $l$

$E_{ij}$  : Evaluation weight % of Measurement  $i$ , submitted by Senator  $j$

$M_{ik}$  : Measurement  $i$ , achieved by the relevant Subnet  $k$

Therefore, the amount of rewards distributed to Subnet  $k$  during epoch  $l$ , will be:

$$\left[ \sum_j \left( \frac{D_j \cdot N_l \cdot E_{ij}}{\sum_j D_j} \cdot \frac{M_{ik}}{\sum_k M_{ik}} \right) \right]$$

Emissions flowing to each subnet are then distributed between Owners, Contributors, and Validators against the Verifiable Robotic Work rules defined by each subnet.

## 5.2 Economic Consumption

There are several areas where the BitRobot Network requires forms of economic consumption. Some of these areas are described below, but we envision more use cases to be defined by the community as the BitRobot Network matures.

### Fee Payment

**Subnet Registration Fees:** These are fees paid per epoch by Subnet Owners to maintain an active subnet. Subnet Registration Fees create an economic cost for joining the BitRobot economy to prevent spam.

**ENT Registration Fees:** These fees are paid per epoch by owners of Robots upon registering new robots into the BitRobot economy.

**Resource Allocation Fees:** These fees are calculated as a small percentage of the revenue paid to Subnet Owners when third parties rent resources via subnets in the BitRobot ecosystem.

**Licensing Fees:** These are fees paid by third parties to commercially license the data or assets stewarded by the BitRobot Foundation.

**Penalties:** These are fees paid by any participant when violating conditions of the BitRobot Network or relevant subnets.

### Collateral

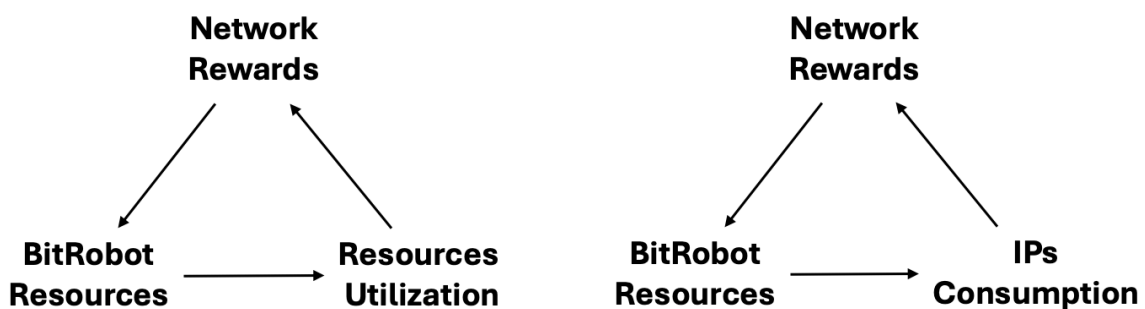
Different components of the BitRobot Network (e.g. ENTs) may require economic collateral to be attached to ensure accountability and enable penalties for malicious actors that might attempt to abuse the BitRobot Network.

### Delegating

BitRobot Network participants may delegate voting power to the BitRobot Senate or Gandalf AI, increasing the weight of the delegatee's votes in determining how BitRobot Network emissions are steered to subnets.

## 5.3 Economic Loops

There are two primary economic loops in the BitRobot Ecosystem:



In the first economic loop, the network rewards are used to acquire useful robotic resources. This can be seen as assembling markets of resources (which may include any useful resource from compute to physical robots to teleoperator man-hours, etc) that can be consumed by any third-party. Thus network utilization incurs a fee, to balance the emissions and utilization.

*Example:* An AI lab hires a fleet of sidewalk robots to evaluate their latest model. The AI lab will pay a resource allocation fee to the network, with the remaining funds being allocated to the hardware owners (via the subnet's definition of Verifiable Robotic Work) whose robots are used in the evaluation of the model.

In the second economic loop, the network rewards are used to both acquire a useful robotic resource, and create a useful output (e.g. dataset or AI model). When that output is licensed, the fees are used to support the BitRobot ecosystem.

*Example:* An AI lab wants to commercially license the robotic arm dataset created by the BitRobot Network. The license fees are used to support the BitRobot ecosystem.

## 6. Future Work

We have presented the BitRobot Network as a powerful framework for creating clusters of resources to accelerate Embodied AI. Below are areas BitRobot Network participants may consider for ongoing improvement:

**Augmenting Verifiable Robotic Work With Cryptographic Proofs:** At launch, the BitRobot Network will focus on Verifiable Robotic Work - a looser constraint (public inputs, public outputs, public transformations) that allows the BitRobot Network to address a wider swath of data collection, processing, and transformation - than what might be economically feasible given the state of zero knowledge protocols today. As the cost curves drop, we envision incorporating cryptographic guarantees to strengthen the economic loops in the ecosystem.

**Markets and Services on top of Network Resources:** We envision more sophisticated interactions being built by BitRobot Network participants on top of the primitives inside of the BitRobot economy. As the BitRobot Network stabilises the capital flows moving through the economy, we envision markets and services to form to improve the capital efficiency for providers on the BitRobot Network.

**Growing the Ecosystem:** BitRobot is maximally useful when it becomes the global hub for talent, manufacturer, service providers, research labs, and AI companies to coordinate their resources against tasks that accelerate Embodied AI. A critical focus will be expanding the scope of participants to accelerate the flywheel of growth in the ecosystem.

**AI Agent Controlled Resources:** A key feature of designing the BitRobot Network to be protocol first is in enabling AI Agents to control physical resources. We expect enabling composability between the BitRobot Network and AI agents to be a continuing focus as the BitRobot Network scales and AI agents mature.

**Evolving Governance:** At launch, we expect the BitRobot Foundation and the Senate to play important roles in helping bootstrap the BitRobot Network and helping steer the flywheel for growth for the BitRobot Network. As the BitRobot Network matures, we expect more responsibilities to transition directly between BitRobot Network participants and Gandalf AI.

## 7. References

1. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. 2017. *Attention is all you need*. <https://arxiv.org/abs/1706.03762>
2. Akash Ghosh, Arkadeep Acharya, Sriparna Saha, Vinija Jain, Aman Chadha. 2024. *Exploring the Frontier of Vision-Language Models: A Survey of Current Methodologies and Future Directions* <https://arxiv.org/abs/2404.07214>
3. Yueen Ma, Zixing Song, Yuzheng Zhuang, Jianye Hao, Irwin King. 2024. *A Survey on Vision-Language-Action Models for Embodied AI* <https://arxiv.org/abs/2405.14093>
4. Julen Urain, Ajay Mandlekar, Yilun Du, Mahi Shafiullah, Danfei Xu, Katerina Fragkiadaki, Georgia Chalvatzaki, Jan Peters. 2024. *Deep Generative Models in Robotics: A Survey on Learning from Multimodal Demonstrations* <https://arxiv.org/abs/2408.04380>
5. Yecheng Jason Ma, William Liang, Hung-Ju Wang, Sam Wang, Yuke Zhu, Linxi Fan, Osbert Bastani, Dinesh Jayaraman. 2024. *DrEureka: Language Model Guided Sim-To-Real Transfer* <https://arxiv.org/abs/2406.01967>
6. Jiayu Xing, Ismail Geles, Yunlong Song, Elie Aljalbout, Davide Scaramuzza. 2024. *Multi-Task Reinforcement Learning for Quadrotors* <https://arxiv.org/abs/2412.12442>
7. Anthony Hu, Lloyd Russell, Hudson Yeo, Zak Murez, George Fedoseev, Alex Kendall, Jamie Shotton, Gianluca Corrado. 2023. *GAIA-1: A Generative World Model for Autonomous Driving* <https://arxiv.org/abs/2309.17080>
8. World Labs. 2024. *Generating worlds*. <https://www.worldlabs.ai/blog>
9. Nvidia. 2025. *Nvidia Cosmos*. <https://www.nvidia.com/en-us/glossary/world-models>
10. Stone Tao, Fanbo Xiang, Arth Shukla, Yuzhe Qin, Xander Hinrichsen, Xiaodi Yuan, Chen Bao, Xinsong Lin, Yulin Liu, Tse-kai Chan, Yuan Gao, Xuanlin Li, Tongzhou Mu, Nan Xiao, Arnav Gurha, Zhiao Huang, Roberto Calandra, Rui Chen, Shan Luo, Hao Su. 2024. *ManiSkill3: GPU Parallelized Robotics Simulation and Rendering for Generalizable Embodied AI* <https://arxiv.org/abs/2410.00425>
11. Zhou Xian, Yiling Qiao, Zhenjia Xu, Tsun-Hsuan Wang, Zhehuan Chen, Juntian Zheng, Ziyang Xiong, Yian Wang, Mingrui Zhang, Pingchuan Ma, Yufei Wang, Zhiyang Dou, Byungchul Kim, Yunsheng Tian, Yipu Chen, Xiaowen Qiu, Chunru Lin, Tairan He, Zilin Si, Yunchu Zhang, Zhanlue Yang, Tiantian Liu, Tianyu Li, Kashu Yamazaki, Hongxin Zhang, Huy Ha, Yu Zhang, Michael Liu, Shaokun Zheng, Zipeng Fu, Qi Wu, Yiran Geng, Feng Chen, Milky, Yuanming Hu, Guanya Shi, Lingjie Liu, Taku Komura, Zackory Erickson, David Held, Minchen Li, Linxi "Jim" Fan, Yuke Zhu, Wojciech Matusik, Dan Gutfreund, Shuran Song, Daniela Rus, Ming Lin, Bo Zhu, Katerina Fragkiadaki, Chuang Gan. 2024. *Genesis: A Generative and Universal Physics Engine for Robotics and Beyond* <https://genesis-embodied-ai.github.io/>
12. Kevin Zakka, Baruch Tabanpour, Qiayuan Liao, Mustafa Haiderbhai, Samuel Holt, Jing Yuan Luo, Arthur Allshire, Erik Frey, Koushil Sreenath, Lueder A. Kahrs, Carmelo Sferrazza, Yuval Tassa, Pieter Abbeel. 2025. *MuJoCo Playground* [https://playground.mujoco.org/assets/playground\\_technical\\_report.pdf](https://playground.mujoco.org/assets/playground_technical_report.pdf)
13. Shikhar Bahl, Russell Mendonca, Lili Chen, Unnat Jain, Deepak Pathak. 2023. *Affordances from Human Videos as a Versatile Representation for Robotics* <https://arxiv.org/abs/2304.08488>
14. Ruijie Zheng, Yongyuan Liang, Shuaiyi Huang, Jianfeng Gao, Hal Daumé III, Andrey Kolobov, Furong Huang, Jianwei Yang. 2024. *TraceVLA: Visual Trace Prompting Enhances Spatial-Temporal Awareness for Generalist Robotic Policies* <https://arxiv.org/abs/2412.10345>

15. Juntao Ren, Priya Sundaesan, Dorsa Sadigh, Sanjiban Choudhury, Jeannette Bohg. 2025. *Motion Tracks: A Unified Representation for Human-Robot Transfer in Few-Shot Imitation Learning*. <https://arxiv.org/abs/2501.06994>
16. Open-X Embodiment authors. 2024. *Open X-Embodiment: Robotic Learning Datasets and RT-X Models*. <https://arxiv.org/abs/2310.08864>
17. Lawrence Yunliang Chen, Kush Hari, Karthik Dharmarajan, Chenfeng Xu, Quan Vuong, Ken Goldberg. 2024. *Mirage: Cross-Embodiment Zero-Shot Policy Transfer with Cross-Painting* <https://arxiv.org/abs/2402.19249>
18. Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Lucy Xiaoyang Shi, James Tanner, Quan Vuong, Anna Walling, Haohuan Wang, Ury Zhilinsky. 2024.  *$\pi 0$ : A Vision-Language-Action Flow Model for General Robot Control* <https://www.physicalintelligence.company/download/pi0.pdf>
19. Kevin Black. 2024. *From Octo to  $\pi 0$ : How to train your generalist robot policy*. <https://www.youtube.com/live/ELUMFpJCUS0?si=3hKJD8C9qpFQokHd&t=16864>