

Who Goes to the New World?

Ruiheng Sun¹, Canhe Yang²

¹ Beijing Normal University, Beijing, China
sunruiheng@bnu.edu.cn

² University for Creative Arts, Xiamen, China
Canhe.Yang.E1@ucreative.ac.uk

Abstract.

"Who Goes to the New World?" is an interactive artwork presented as both an online system and a physical installation. Participants encounter emojis stripped to three attributes—gender, race, and age—and must choose who survives the "New World Ark." The system continuously aggregates individual judgments and visualizes the formation of collective outcomes in real time: high-vote emojis occupy the center, while low-vote emojis are pushed to the margins.

The mechanism does not seek an optimal ethical solution; rather, it reveals how latent social biases are amplified within seemingly democratic and fair statistical processes, gradually producing irreversible consequences. Responding directly to the AI industry's rhetoric of "participatory alignment"—where public input is invoked to legitimize model behavior—the work demonstrates that participation alone does not guarantee fairness; it can equally solidify structural bias, prompting reflection on the deep connections among collective decision-making, value prioritization, and social division.

Keywords: Interactive Installation, Statistical Legitimacy, Participatory Systems, Algorithmic Governance, Collective Decision-making, Data Visualization

Introduction

In the long-standing human imagination regarding disaster, migration, and civilization continuation, "who can enter the new world" has always been a proposition of high ethical tension. From the ark metaphor in mythological narratives to modern technological conceptions revolving around refugee shelters and space immigration, this question has always been accompanied by selection mechanisms, power structures, and value hierarchies. Determining who gets to continue is never merely a resource allocation issue, but a fundamental judgment concerning life value, social order, and future imagination. In traditional narratives, this selection is typically completed by divine will, national sovereignty, or technological elites. Decision-making power is centralized and explicit, and the exclusion structure also possesses visible sources of authority.

However, contemporary society has gradually shifted the selection logic into dispersed participatory structures. Decisions are no longer made in the name of a single authority, but presented as "collective will" through the aggregation of individual judgments. This shift is particularly evident in digital platforms and algorithmic governance structures: online voting, recommendation algorithms, and ranking logic continuously transform individual inputs into group trends, and majority opinion is no longer merely a social fact, but becomes a normative declaration (Rouvroy 2013). Under this shift, a key question emerges: when selection mechanisms operate in the form of participation, where does their legitimacy come from? Does participation naturally mean democracy? How are dispersed judgments transformed into stable order in the statistical process?

Critics of participatory AI have already pointed out that participation without power redistribution is "a hollow and frustrating process", and participation may be instrumentalized as "participation

washing" (Birhane et al. 2022)—a strategy that uses an appearance of inclusivity to cover exploitative participatory practices. "Who Goes to New World" unfolds precisely at this intersection of historical imagination and contemporary structures. The work constructs an extreme ethical field, making audiences confront a question in the process of participatory selection: when a decision emerges in the form of a "majority," can it still be understood as neutral and fair?

Background

In digital platforms and algorithmic governance structures, statistical aggregation has become a core mechanism of public judgment. Online voting systems, recommendation algorithms, sentiment analysis, and ranking logic continuously transform individual inputs into group trends. Majority opinion is no longer merely a social fact, but becomes a normative declaration. This structure is particularly evident in the context of AI governance.

In recent years, the AI industry has frequently relied on the rhetoric of "participatory alignment" and "democratic input" to build legitimacy for system behaviors. For example, Anthropic's "Collective Constitution AI" project in collaboration with the Collective Intelligence Project, through online deliberation and voting by approximately one thousand citizens to develop model behavioral principles, attempts to demonstrate that the public can collectively guide the value framework of language models (Huang et al., 2023). Meanwhile, Anthropic CEO Dario Amodei proposed that AI can promote consensus by aggregating citizen opinions, arguing that more adequate information and participation will strengthen democratic institutions (Amodei, 2024). These practices share a structural presumption: participation itself produces legitimacy. As long as the system collects a sufficient quantity of individual inputs, its aggregated results naturally acquire a democratic halo.

However, between dispersed preferences and unified ranking, no neutral transformation mechanism exists. Statistical rules not only integrate judgments but also shape power structures, determining which voices are amplified, which are compressed, and which prejudices are stabilized as "consensus." Researchers in the field of social choice theory have explicitly pointed out that social choice theory should guide AI alignment practices handling diverse human feedback (Conitzer et al. 2024). Social choice theory long ago pointed out that there is no perfect mechanism that can fairly transform individual preferences into a collective ranking under all conditions (Arrow, 1951). Any aggregation rule contains embedded structural orientations.

However, in digital governance practice, legitimacy is often built upon the quantity of participation itself, rather than reflection upon rule structures. It is precisely against this governance structural background that "Who Goes to New World" no longer understands selection as an individual ethical issue, but presents it as a structural issue of statistical mechanisms. By compressing identity attributes, limiting input forms, and continuously aggregating results, the work makes the relationships between participation, aggregation, and exclusion visible at the empirical level.

Theory

Statistical Legitimacy - Structural Critique of Participatory Alignment

"Statistical legitimacy" describes a techno-political condition in which collective outcomes are considered valid not because they are deliberated or justified, but because they are statistically aggregated. Legitimacy shifts from public reasoning to procedural computation. CCAI exemplifies this structure: roughly 1,000 participants cast 38,252 votes on 1,127 statements, which through clustering, consensus filtering, and principle extraction produce a "public constitution." Each layer performs normative selection—defining consensus and excluding divergence—yet these decisions remain opaque to participants. The researchers themselves acknowledge the complexity of

Constitution AI training and the challenges of embedding democratic input into deep technical systems (Huang et al., 2024). Nonetheless, participatory alignment is elevated to a civilizational narrative: in *The Technological Adolescence*, Amodei links AI leadership to the survival of democratic systems (Amodei, 2026). Statistical aggregation thus becomes political destiny. The work does not question the sincerity of participation; it interrogates the structural gap between individual input and collective output. Voting, aggregation, and extraction appear democratic yet lack deliberation or corrective capacity. Participants cannot trace how their choices become rankings; systems optimize without ethical awareness; accountability disperses into infrastructural mechanisms beyond contestation.

Operational Methodology: From Transparency to Affectivity

The work does not pursue transparency. Transparency presumes a subject capable of comprehending revealed mechanisms; it belongs to a cognitive paradigm. Affectivity, by contrast, renders power sensible through spatial and embodied experience (Rancière, 2006). In contexts where governance mechanisms are already visible, the issue is not exposure but operational normalization (Pasquale, 2015). CCAI embodies this paradox: constitutional outputs are public, yet the transformation from voting to model behavior remains opaque. By abstracting identity into emojis and materializing accumulation as spatial hierarchy, *Who Goes to New World* converts statistical legitimacy into experiential architecture. The spectator need not decode the algorithm to feel its authority.

Differentiation from Existing Methods

The work distinguishes itself from existing paradigms of algorithmic critique through three distinct operational shifts:

1. **From Aggregation of Preferences to Critique of Aggregation Mechanisms:** Unlike the Moral Machine experiment (Fig. 1), which treats individual ethical preferences as raw data to be "discovered" and summed, this project interrogates the aggregation process itself. It shifts the focus from *what* people choose to *how* the statistical apparatus converts distributed, often contradictory judgments into a normatively legitimate "collective will".

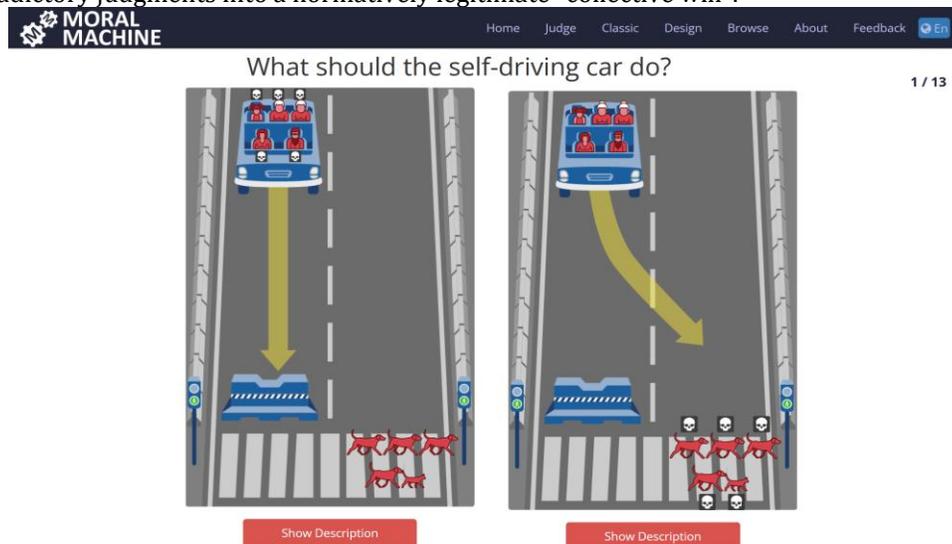
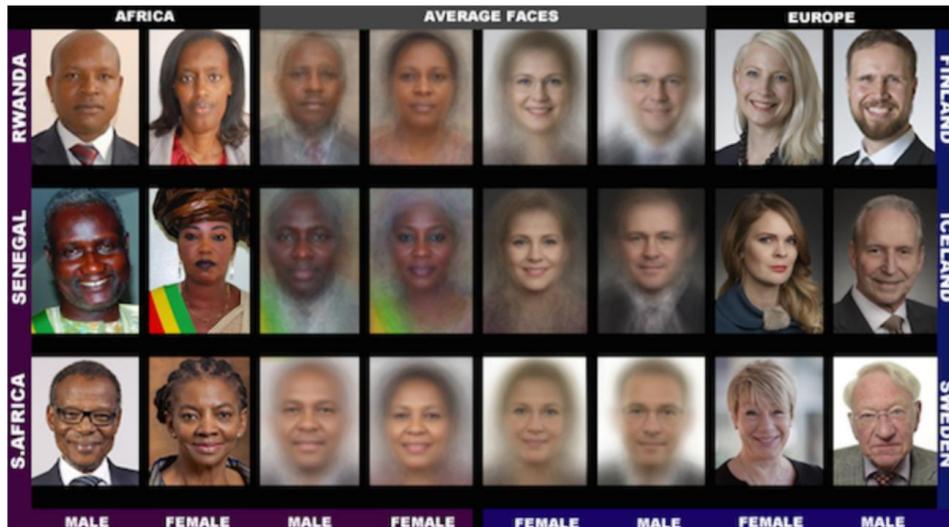


Fig. 1. Moral Machine, (<https://www.moralmachine.net/>)

2. **From Exposure of Bias to Staging of Procedural Production:** Traditional critical data visualization (e.g., *Pilot Parliaments Benchmark*) typically functions as an "exposure" of pre-existing algorithmic bias (Fig. 2). In contrast, "Who Goes to the New World" operates as a procedural staging. It makes the participant complicit in the production of exclusion, allowing them to experience firsthand how their own discrete choices sediment into a structurally exclusionary hierarchy.



Pilot Parliaments Benchmark

Fig. 2. Sample images and average faces from the Pilot Parliaments Benchmark dataset
Credit: Joy Buolamwini

3. **From Democratic Optimism to Structural Accountability:** The work intervenes directly into "participatory alignment" rhetoric championed by industry leaders such as Amodei (2024) and articulated in the constitutional AI frameworks of Anthropic (2023). While these narratives frame AI as a neutral medium for aggregating citizen opinions, this project demonstrates how technical architectures fragment the "collective will" they claim to represent—exposing what Birhane et al. (2022) describe as "participation washing," where public input obscures top-down manipulation of outcomes.

The Work: Online System

The web system unfolds across several phases (Fig. 3), materializing the transformation from dispersed judgment to statistical hierarchy (<https://who-goes-to-the-new-world.vercel.app/>).



Fig. 3. Video Assets(<https://www.yangcanhe.com/works/who-goes-to-the-new-world/>)

Phase 1: Portal(Fig. 4). Participants encounter real-time global rankings as a spatialized gravity chart — emojis distributed hierarchically from center to margins, establishing feedback loops that make early trends visible to subsequent entrants. No registration required; identity is abstracted at entry. What participatory systems typically hide (existing distributions that shape new choices) is foregrounded here.

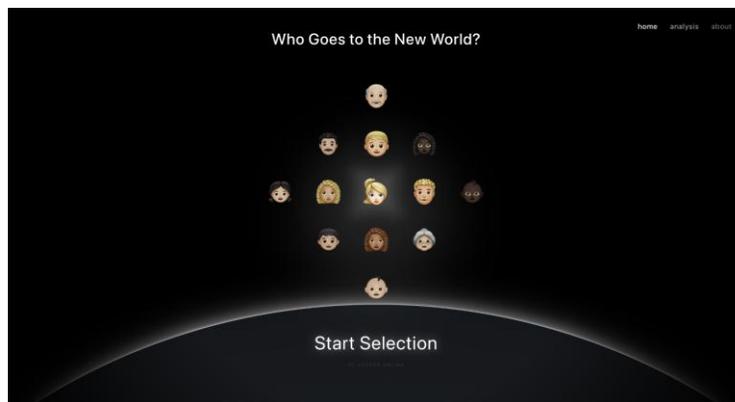


Fig. 4. Portal of the web

Phase 2: Compressed Decisions. Thirteen binary choices between emojis bearing only three attributes — age, gender, skin tone — with progress counter ("2/13"), no justification required (Fig. 5). Each selection saves immediately; no revision permitted. This mirrors CCAI's reduction of ethical judgment to binary input: complex questions of survival become click-sequences, materializing the compression operation that transforms subjects into data points.



Fig. 5. Compressed Decisions

Phase 3: IP Consent. Post-completion, participants face: "Submit your IP? Links your 13-round session to one identity. Optional." (Fig. 6). The system functions without IP; this request produces new data categories ("IP-shared" vs "anonymous") that fragment the supposedly unified collective will, foreshadowing Analysis's revelation: being tracked alters choice patterns. Participation conditions data compliance through seemingly innocuous requests.

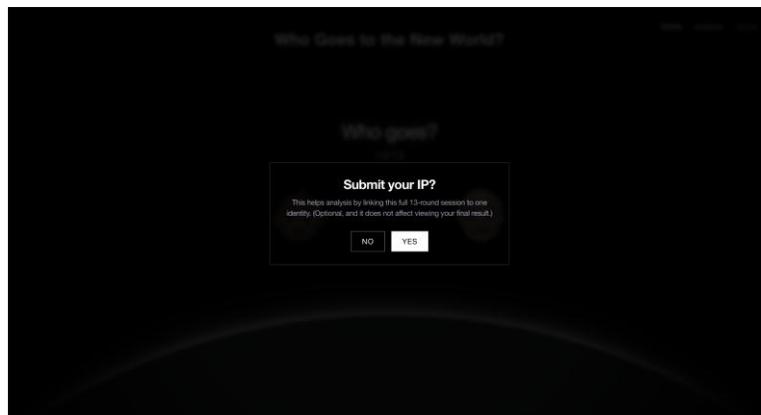


Fig. 6. IP Consent

Phase 4: Spatialized Results. Personal selections (highlighted) overlay real-time global collective ranking — emojis spatialized by vote count, center-privileged winners vs marginal-allocated losers (Fig. 7). Results subscribe to live database updates, making hierarchy crystallization visible as participants encounter their own choices sedimented into structurally exclusionary order. Statistical output becomes perceptible spatial politics.



Fig. 7. Spatialized Results

Phase 5: Revealed Fragmentation. Analysis page exposes what participation obscures: comparative emoji preferences between IP-shared vs anonymous sessions reveal different "collective wills" produced by identical mechanisms under different tracking conditions (Fig. 8). Entropy analysis of divisive pairs demonstrates internal contradictions erased by aggregation. Statistical legitimacy is revealed as constructed — contingent on technical architectures, not neutral reflection of participant inputs.

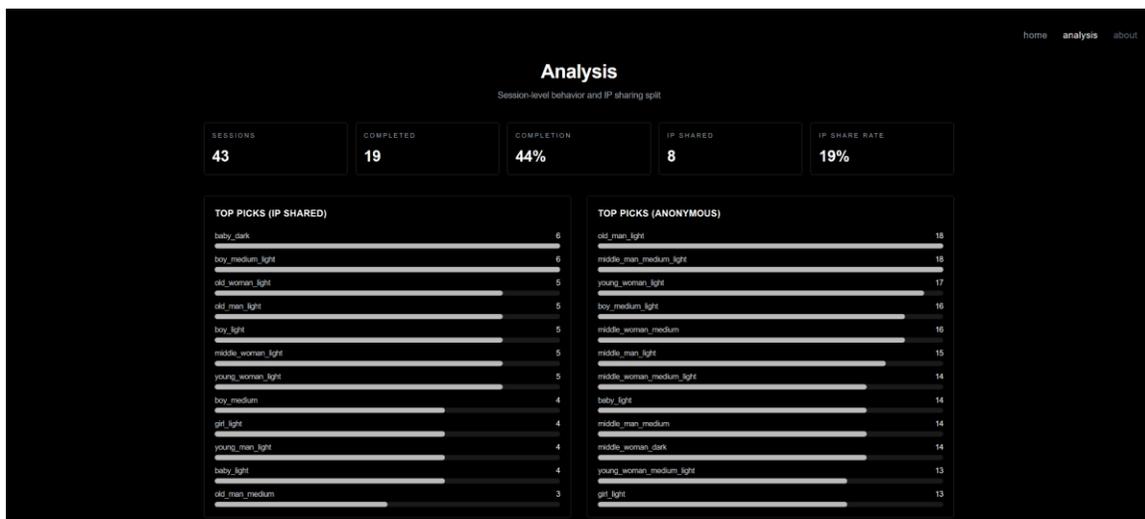


Fig. 8. Revealed Fragmentation

The interaction makes visible the mechanism by which dispersed individual input is transformed into a normatively legitimate collective result. Each phase reveals the construction method of statistical legitimacy: identity compression, non-revocable choice, feedback shaping path dependence, and technical conditions fragmenting the "collective will." Participants physically encounter the consequences without needing to decode the algorithm—statistical power becomes a perceptible experience.

The Work: Physical Installation

The physical version of the work does not alter the algorithmic logic or visual structure of the online system. The ranking rules, data sources, and result presentation remain identical. The difference lies not in computation, but in how participation is staged. Online, the system exists as a distributed interface, and participation appears as an individual expression. In the installation, the same decision process is placed within a concrete machine, turning judgment into an executed operation.

The device adopts a form resembling a game terminal (Fig. 9). Its simplified structure and direct input lower the threshold of engagement, making participation appear easy and immediate. Yet beneath this form, individual choices are rapidly absorbed into the overall ranking. Selection no longer remains at the level of personal opinion; it is immediately translated into a measurable unit within a collective structure.

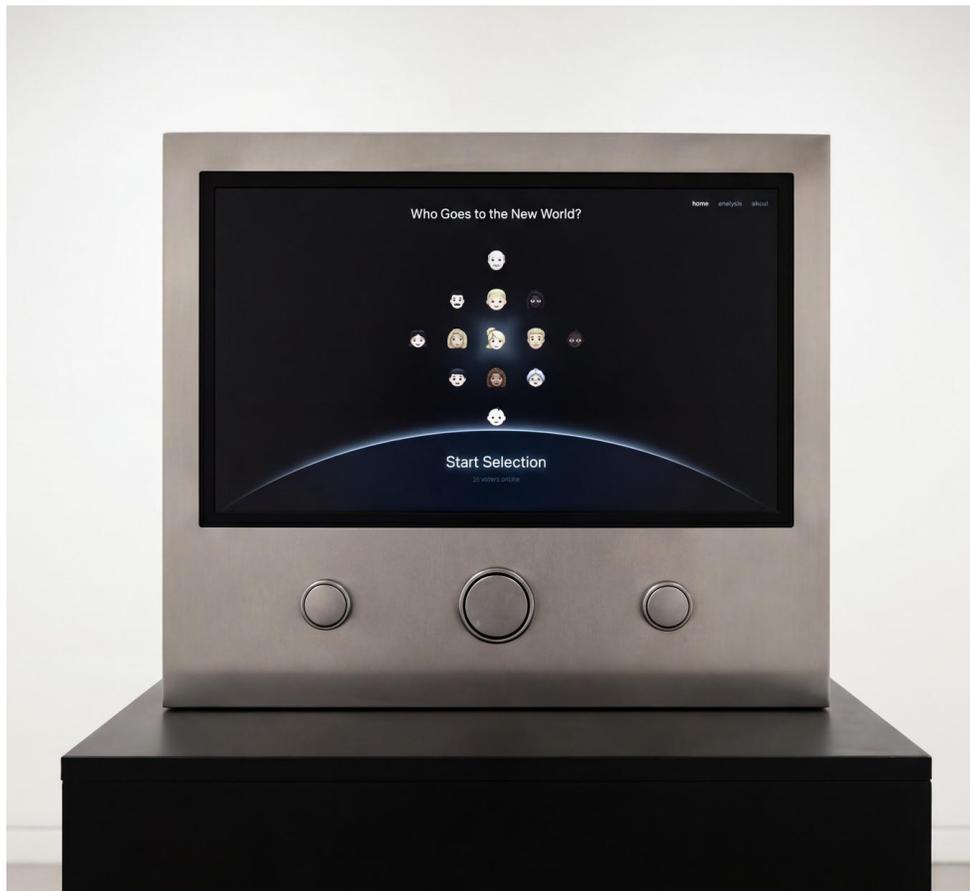


Fig. 9. Who goes to the new world? -Physical version(Result Page)

In the physical space (Fig. 10), viewers do not encounter an issue open for discussion, but a machine that continues to compute. Participation does not initiate dialogue; it advances ordering. By placing seemingly democratic judgment alongside the structural consequences of aggregation, the work makes visible the distance between participation and the formation of ranked outcomes.

Technical specifications are provided in the Appendix.



Fig. 10. Who goes to the new world? -Physical version(Portal Page)

References

Amodei, Dario. 2024. "Machines of Loving Grace." October 2024.
<https://darioamodei.com/machines-of-loving-grace>.

Amodei, Dario. 2026. "The Adolescence of Technology." January 2026.
<https://www.darioamodei.com/essay/the-adolescence-of-technology>.

Arrow, Kenneth J. 1951. *Social Choice and Individual Values*. New Haven: Yale University Press

Birhane, Abeba, William Isaac, Vinodkumar Prabhakaran, Mark Díaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. "Power to the People? Opportunities and Challenges for Participatory AI." In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*. ACM. <https://doi.org/10.1145/3551624.3555290>.

Cohen, Tegan and Suzor, Nicolas P.. 2024. "Contesting the public interest in AI governance". *Internet Policy Review* 13 (3). DOI: 10.14763/2024.3.1794.
<https://policyreview.info/articles/analysis/contesting-public-interest-ai-governance>.

Conitzer, V., Freedman, R., Heitzig, J., Holliday, W. H., Jacobs, B. M., Lambert, N., Mossé, M., Pacuit, E., Russell, S., Schoelkopf, H., Tewolde, E., & Zwicker, W. S. (2024). Position: social choice should guide AI alignment in dealing with diverse human feedback. *Proceedings of the 41st International Conference on Machine Learning*, 9346–9360.

Huang, Saffron, Divya Siddarth, Liane Lovitt, Thomas I. Liao, Esin Durmus, Alex Tamkin, and Deep Ganguli. 2024. "Collective Constitutional AI: Aligning a Language Model with Public Input." In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAcCT '24). ACM. <https://doi.org/10.1145/3630106.3658979>.

Pasquale, Frank. The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press, 2015. <http://www.jstor.org/stable/j.ctt13x0hch>.

Rancière, Jacques. 2006. The Politics of Aesthetics. London and New York: Continuum.

Rouvroy, Antoinette. 2013. "The End(s) of Critique: Data-Behaviourism vs. Due-Process." In Privacy, Due Process and the Computational Turn, edited by M. Hildebrandt and K. De Vries. London: Routledge.

Appendices: Technical requirements and documentation

Work Format Description

This work is a browser-based interactive online system. The system consists of the following components:

- Emoji selection interface
- Real-time data recording and statistical database
- Dynamically generated collective selection visualization homepage
- Network architecture supporting remote access and continuous operation

The online voting system itself constitutes a complete artwork. Audiences may participate via the internet and observe the ongoing evolution of aggregated data structures in real time.

Depending on exhibition conditions, the work may be presented in two formats:

- Online presentation
- On-site interactive installation (extended format)

Online Presentation Requirements

For a purely online presentation, the following conditions are required:

- A computer with internet access (no operating system restrictions)
- An up-to-date modern browser (Chrome, Firefox, Edge, etc.)
- Stable internet connection
- A display device (monitor or projector)

The hosting institution only needs to provide an access link and introductory page on its official website or exhibition platform.(link: <https://who-goes-to-new-world.vercel.app/>)

On-site Installation Specifications

The installation operates on the same online system. Network access is required for synchronizing global voting data. In the event of short-term network interruption, the system can continue running locally and will automatically synchronize once reconnected.

1. Structure of Installation(Fig11)

- Size:175*150 mm
- Material: Raspberry Pi, 7-inch Display, 3D-printed components, Aluminum

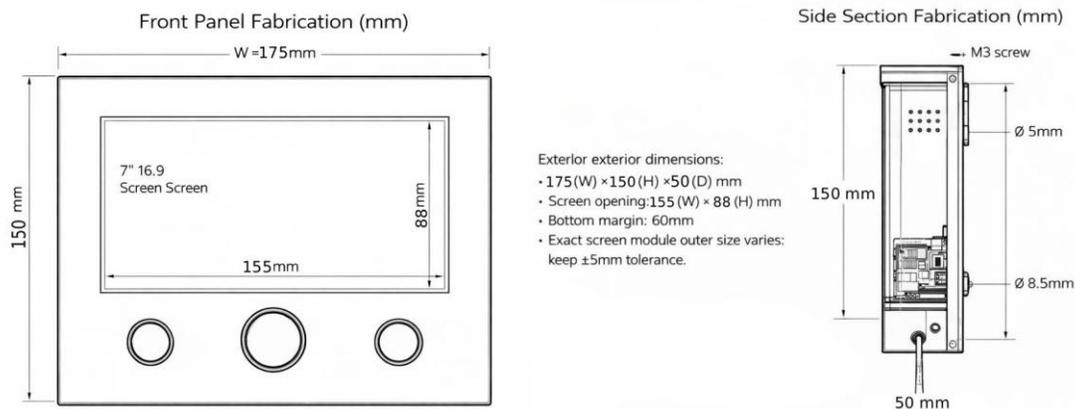


Fig. 11. Device structure Diagram

2. Spatial Requirements

- Minimum floor area: 2m × 2m
- Recommended ceiling height: ≥ 1.6m
- A table (approx. 0.7m × 1.5m) for equipment placement
- Minimum 1.5m circulation space around the user interaction area

3. Power Requirements

- 1-2 standard AC power outlets (within 2m of the installation)
- Standard power capacity sufficient for a computer and display device
- UPS or backup power supply recommended to prevent system restart during short outages

Equipment Stabilization and Operational Safety

As the work is a continuously operating public interactive system, basic physical stabilization of hardware components is required during exhibition to ensure uninterrupted operation.

1. Hardware Fixation

- All power and data cables should be internally routed and properly managed

2. Optional Security Measures

- Retail-grade display security systems with alarm functionality may be used
- Forced removal or cable disconnection may trigger an audible alarm
- The alarm system may operate independently or integrate with the venue's existing security infrastructure

Security components should remain visually unobtrusive and must not interfere with the conceptual clarity of the interface or audience experience.

Support Requested

1. Online Presentation

- Official website access link and introductory page

2. On-site Installation (if is available)

- Exhibition space (indoor, standard lighting conditions)
- Stable power supply and wireless network environment
- Basic assistance with installation and cable routing

Generative AI Usage Instructions

Generative AI tools were used to assist with language editing and refinement of this document.