

PANIC!

BEN SWIFT, School of Cybernetics, Australian National University, Australia

PANIC! (Playground AI Network for Interactive Creativity) is an interactive installation that explores the behaviour of *connected* AI models. Viewers enter text prompts which are transformed as text/audio/images through a "network" of generative AI models. Each output becomes the input for the next iteration, creating an endless cycle of AI-mediated transformation.

Additional Key Words and Phrases: Installation, Generative AI, text-to-music, music-to-text, Feedback Systems

1 Program Notes

PANIC! is an interactive installation where *you* can play with feedback loops of generative AI models hooked end-to-end. Instead of one-and-done interactions with these models (producing a single text, image or audio output) PANIC! explores how different ways of connecting these models up can give rise to different patterns of outputs, emergent behaviours, recurring patterns, and degenerate cases. Today, the low barrier to entry in generative AI model platforms means that more and more people of us are using them. But just because we can put our text/images/audio into these models and have them provide new text/images/audio in return, does it mean that we should?

2 Project Description

PANIC! explores the musical and sonic serendipity that emerges through interconnected networks of generative models.

Visitors interact with PANIC! by typing a text prompt into a terminal which is physically located in the installation space. This prompt is fed into a chain of generative AI transformations, using various cloud-hosted AI platforms including Google, OpenAI, and Replicate. As each output becomes the input for the next iteration, unexpected musical motifs, sonic textures and imagery emerge through the feedback loop, creating an ever-evolving soundscape that fills the installation space. This feedback loop is the core of PANIC!'s generative process, and is shown in Figure 1.

PANIC! invites visitors to observe how their text inputs shape the evolution of the generative audio over time, while also highlighting how the system develops its own patterns and tendencies independent of human intervention. The multiple speakers and displays arranged throughout the space create an immersive sonic environment, allowing visitors to experience how different configurations of AI model networks can give rise to varying musical behaviors and emergent sonic properties. PANIC! functions as both an interactive musical instrument and a platform for exploring questions about human and machine agency in creative sound generation. The endless cycle of transformation between linguistic, musical and visual expression provides a lens for examining how meaning and musicality are negotiated in closed-loop AI systems.

3 Technical Notes

While the generative AI models themselves (and the open source application server[7] that powers the PANIC! installation) run in the cloud, their outputs are displayed/played back "live" in the installation space via a heterogeneous array of displays and loudspeakers as shown in Figure 2 (an example screenshot is shown in Figure 3). This functionality is web-based, so any internet-connected screen or loudspeaker can be part of the PANIC! installation. For example, a flat-panel TV can be added into the system by adding a Raspberry Pi to run the web browser, sending audio/video to the display via HDMI.

The physical configuration of PANIC! is therefore highly flexible: it is currently installed in the Birch Building at the Australian National University (the NIME 2025 host venue), but will be re-deployed at the NIME 2025 venue in a site-specific configuration according to space and artistic considerations.

PANIC! works best when configured across multiple physical displays and loudspeakers—see Figure 4 for an example setup with 2x flat-panel TV displays for images and audio, and 2x Vestaboard[9] split-flap displays for text output.

3.1 Generative AI Model list

The PANIC! system is configurable and can call on a variety of generative AI models, each performing a specific transformation within the feedback loop. These models are hosted on various cloud platforms and are chosen for their capabilities

Author's Contact Information: Ben Swift, School of Cybernetics, Australian National University, Canberra, Australia.



This work is licensed under a Creative Commons Attribution 4.0 International License.

NIME '25, June 24–27, 2025, Canberra, Australia

© 2025 Copyright held by the owner/author(s).

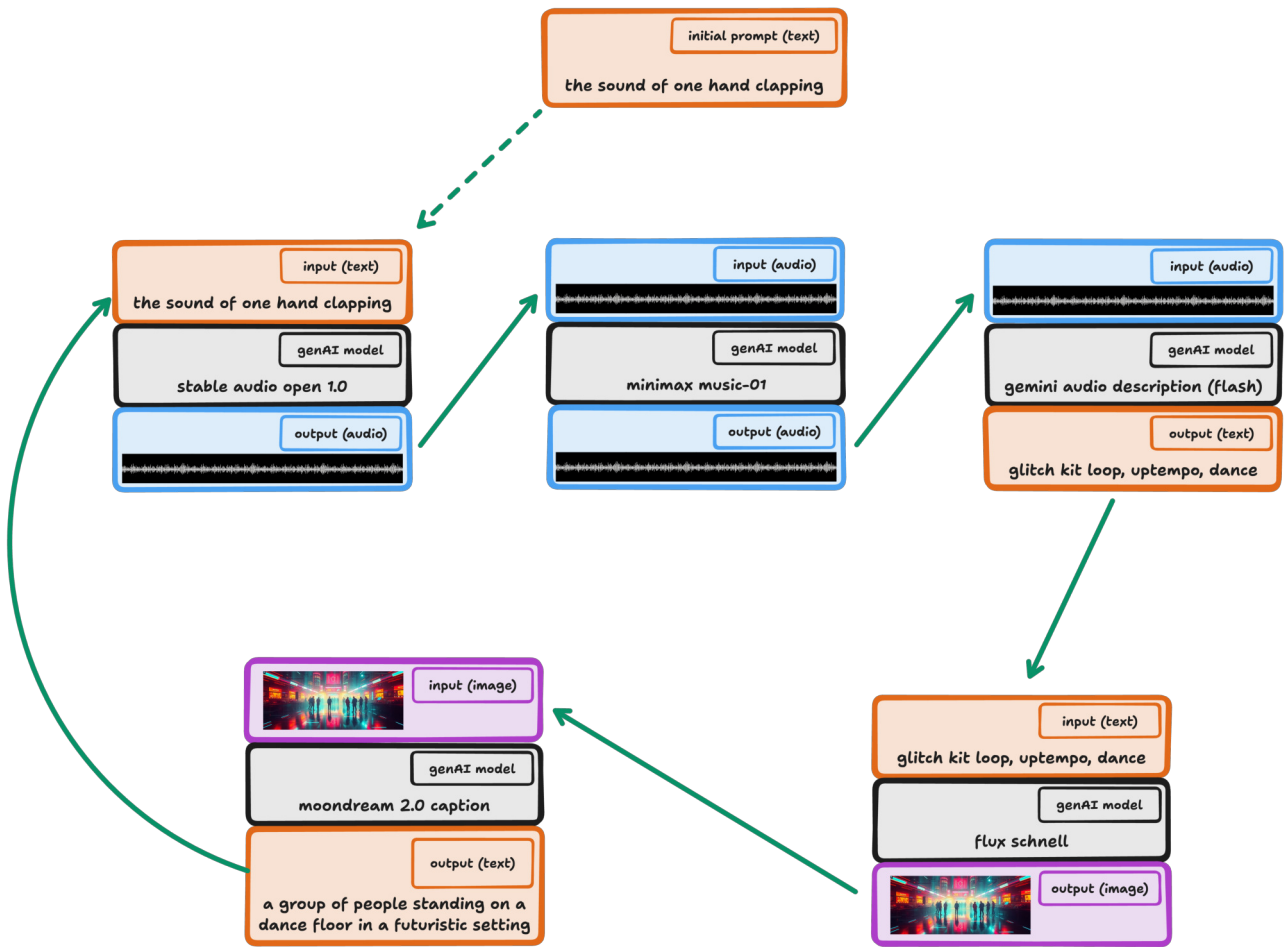


Fig. 1. Logical information flow within the PANIC! system. Once the viewer starts the loop running with an initial prompt, the system will continue to pass the newly-generated text/audio/images through in a closed feedback loop (until the next new prompt comes in). Note that this figure shows one particular “network” of generative models; PANIC! can be easily re-configured to use a different network as long as the output-input modalities match at each link in the chain. The current PANIC! model list is given in Section 3.1, although PANIC! is regularly updated as new models are released.

in specific modalities. The models currently integrated into the system are listed below, grouped by their input and output types:

- **Text-to-Audio:** Models that generate audio based on a textual description.
 - Meta’s MusicGen [1]
 - Meta’s MAGNeT [10]
 - Stability AI’s Stable Audio Open [2]
- **Audio-to-Audio:** Models that generate audio based on a textual description.
 - MiniMax’s Music-01 [6]
- **Audio-to-Text:** Models that transcribe or describe audio content into text.
 - Google’s Gemini [8]
- **Text-to-Image:** Models that generate images from text prompts.
 - Black Forest Labs’ FLUX.Schnell [4]
- **Image-to-Text:** Models that generate textual descriptions (captions) for images.
 - Moondream 2 [3]
 - BLIP-2 [5]

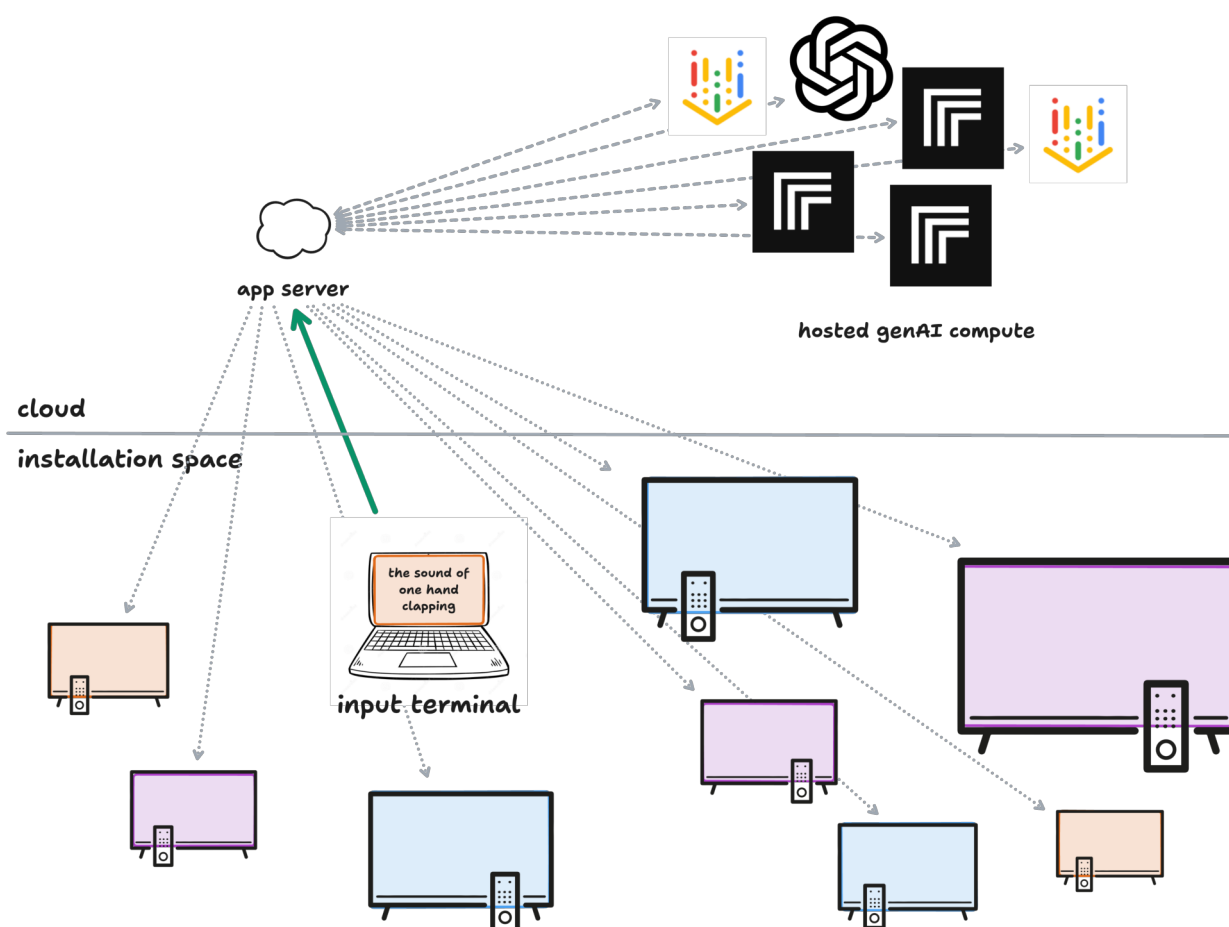


Fig. 2. The physical architecture of the PANIC! system. The configuration of the displays and loudspeakers in the installation space is site-specific, based on both technical and artistic considerations.

4 Ethical Standards

PANIC! was developed at the ANU School of Cybernetics with funding from the Australian National University. The installation uses commercial AI models in accordance with the respective platform terms and conditions. Viewer text inputs (prompts) are processed through these models, but no other identifying information about the viewer is collected or stored. While the system includes content filtering, we recommend adult supervision for children interacting with the installation due to the unpredictable nature of AI-generated content.

The work aims to provoke discussion about the ethical implications of AI art tools, including questions of bias, agency, and the relationship between human and machine creativity. Installation power usage and computational resources have been considered and optimized where possible for environmental sustainability.

References

- [1] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. 2024. Simple and Controllable Music Generation. arXiv:2306.05284 [cs.SD] <https://arxiv.org/abs/2306.05284>
- [2] Zach Evans, Julian D. Parker, CJ Carr, Zack Zukowski, Josiah Taylor, and Jordi Pons. 2025. Stable Audio Open. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1–5. <https://doi.org/10.1109/ICASSP49660.2025.10888461>
- [3] Vikhyat Korrapati. 2024. Moondream 2. GitHub repository. <https://github.com/vikhyat/moondream>
- [4] Black Forest Labs. 2024. FLUX. <https://github.com/black-forest-labs/flux>.
- [5] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models. <https://doi.org/10.48550/ARXIV.2301.12597>
- [6] MiniMax. 2024. Music-01. Web. <https://www.minimax.io/news/music-01>
- [7] Ben Swift. 2022. PANIC!: Playground AI Network for Interactive Creativity. GitHub repository. <https://github.com/ANUcybernetics/panic/>

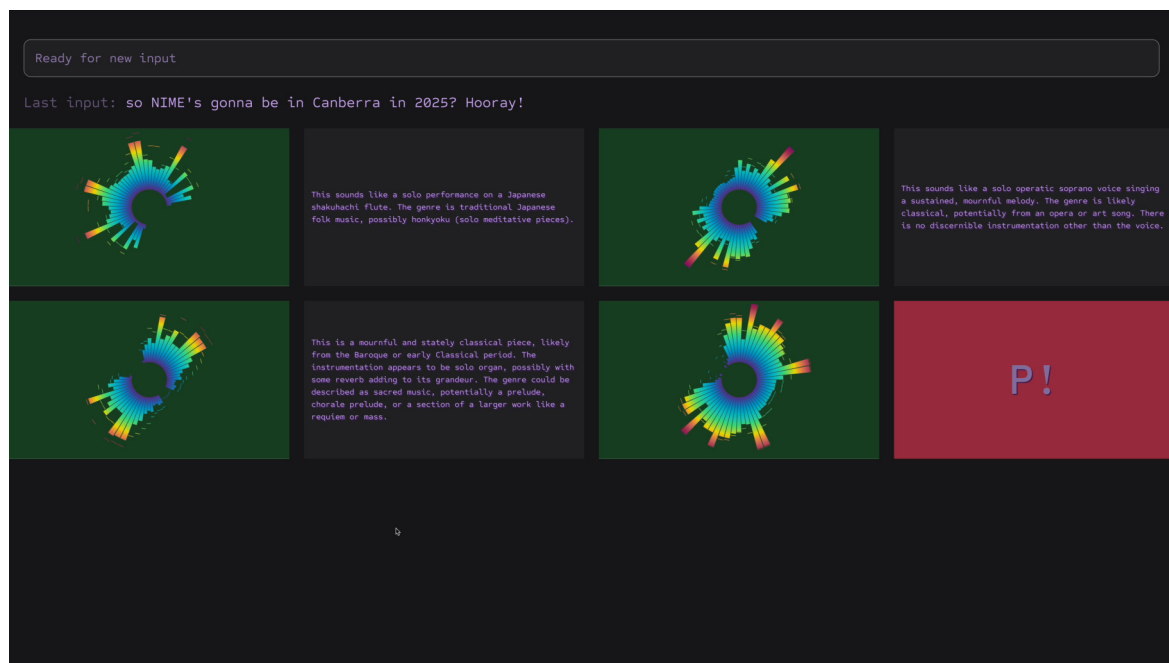


Fig. 3. A screenshot from one of the PANIC! displays showing the text and audio outputs of successive models in a text-audio-text network.



Fig. 4. A PANIC! installation configuration from SXSW Sydney 2024—2x flat-panel displays, 2x Vestaboard split-flap displays for text output.

- [8] Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, et al. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530* (2024). <https://arxiv.org/abs/2403.05530>
- [9] Vestaboard, Inc. 2024. Vestaboard. Web. <https://www.vestaboard.com>
- [10] Alon Ziv, Itai Gat, Gael Le Lan, Tal Remez, Felix Kreuk, Alexandre Défossez, Jade Copet, Gabriel Synnaeve, and Yossi Adi. 2024. Masked Audio Generation using a Single Non-Autoregressive Transformer. *arXiv:2401.04577* [cs.SD] <https://arxiv.org/abs/2401.04577>