How Foundation Models Are Reshaping Non-Invasive Brain–Computer Interfaces: A Case for Novel Human Expression and Alignment

Albert Barqué-Duran, Ada M. Llauradó Crespo
IASlab (Interactive Arts Science Laboratory)
Digital Engineering School (ETSELS)
La Salle Campus Barcelona - Universitat Ramon Llull (Barcelona, Spain)
albert.barque@salle.url.edu

Abstract

SYNAPTICON is a research prototype at the intersection of neuro-hacking, non-invasive brain-computer interfaces (BCIs), and foundation models, probing new territories of human expression, neuroaesthetics, and AI alignment. Envisioning a cognitive "Panopticon" where biological and advanced synthetic intelligent systems converge, it enables a pipeline that couples temporal neural dynamics with pre-trained language models and operationalizes them in a closed loop for expression. At its core lies a live "Brain Waves-to-Natural Language-to-Aesthetics" system that translates neural states (i.e. electroencephalography (EEG)) into decoded speech, and then into immersive audiovisual output and content; shaping altered perceptual experiences and inviting audiences to directly engage with the user's mind. *SYNAPTICON* provides a reproducible reference for foundation-model-assisted BCIs, suitable for advanced studies of human–machine interaction (HMI).

1 Introduction: from Signal to Semantics

1.1 Foundation Models "meet" the Brain

Non-invasive brain-computer interfaces (BCIs) are rapidly evolving—from letter-by-letter spellers to systems capable of fluent language and detailed imagery. Foundation models now decode continuous inner speech from functional magnetic resonance imaging (fMRI) (Tang et al. [2023]; Ferrante et al. [2023]) and reconstruct visual scenes using diffusion pipelines (Özçelik and VanRullen [2023]), exponentially expanding BCI bandwidth. As large language models (LLMs) complete neural prompts, the extended mind thesis (Clark and Chalmers [1998]) materialises: cognition becomes a hybrid of brain and model. This raises critical questions around AI alignment, mental privacy and neuro-rights—cognitive liberty, mental integrity, and psychological continuity (Ienca and Andorno [2017]).

One line of recent research focuses on translating electroencephalography (EEG) signals into natural language. Early EEG-based BCI communication systems relied on explicit selection paradigms (e.g. P300 spellers) to spell out messages, but new approaches attempt direct decoding of imagined speech from brain activity. Despite the low signal-to-noise ratio of EEG, steady progress has been made in mapping neural signals to text [Lopez-Bernal et al., 2022]. Modern deep learning techniques enable open-vocabulary EEG-to-text translation by coupling neural feature encoders with pre-trained language decoders [Murad et al., 2025]. For example, Murad et al. [2025] used a long short-term memory (LSTM) encoder with a bidirectional and auto-regressive transformer (BART) to achieve state-of-the-art EEG narrative decoding. Others incorporate representation learning to improve performance: Lu et al. [2024] leveraged a contrastive EEG–text pretraining strategy, and Duan et al. [2023] introduced a discrete vector quantization tokenization ("DeWave") to better align

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Creative AI Track.

EEG segments with language model inputs. These advances demonstrate the feasibility of directly translating non-invasive brain activity into text, though accuracy remains limited.

In parallel, large foundation models have been applied to decode language from other neural modalities. Tang et al. [2023] reconstructed continuous language from fMRI by aligning brain activity with a generative language model, effectively translating thoughts to text. Ferrante et al. [2023] similarly used a multimodal diffusion model to recover viewed images and descriptive text from fMRI data, illustrating how brain signals can be embedded into the rich latent space of generative models. Such studies represent a paradigm shift: rather than training task-specific decoders from scratch, brain representations are mapped into pretrained model spaces—leveraging vast prior knowledge to greatly expand BCI bandwidth and expressiveness.

1.2 Creative BCIs for Artistic Research and Practice

A growing body of work also merges BCIs with generative art. Artists and researchers have long used neural signals as a medium for creative expression—from early EEG-driven music performances to "brain painting" interfaces for paralyzed patients. Contemporary BCI art installations visualize or sonify neural activity in real time, turning cognitive states into dynamic audiovisual experiences [Gürkök and Nijholt, 2013, Nijholt, 2019]. For example, Schreiner et al. [2025] developed an EEG-driven wearable system that modulates an interactive dress's visuals and motions based on the wearer's mental engagement. The emergence of "brain art" shows that brain data can directly drive generative media, blurring boundaries between neuroscience and creative practice. This project aims to go a step further and showcase an applied and novel case of human-machine interaction (HMI).

1.3 Closed-Loop AI Multimodal Generation from Brain Signals

Against this backdrop, *SYNAPTICON* operates as a live, generative EEG-to-text-to-aesthetics pipeline that pushes BCI into the artistic realm (please see Figure 1). It leverages an OpenBCI 16-channel headset and a hybrid LSTM encoder with a pre-trained BART decoder (a language foundation model) to translate brainwaves into text in real time. That text is immediately fed into multimodal generative AI models to produce synchronized audiovisual outputs and/or orchestrate a full audiovisual live-show. To our knowledge, this is the first demonstration of an online imagined-speech BCI driving live artistic content. Unlike prior EEG decoding systems—which were typically offline and task-oriented—*SYNAPTICON* functions as a performance piece, transforming the artist's neural activity into an evolving narrative and visual experience on stage. This one-of-a-kind configuration exemplifies a new genre of human–AI co-creativity, where a human's inner thoughts seamlessly trigger an artistic piece/experience.

2 *SYNAPTICON*: an EEG-to-Text-to-Aesthetics framework

2.1 Data Collection:

The user (i.e. the performer) wore an OpenBCI's Ultracortex Mark IV EEG headset (Cyton-Daisy 16-channel) while being presented with a series of sentences visually (i.e. on a screen). Each sentence was displayed for a fixed duration (10 seconds), during which EEG data was recorded and labeled with that sentence's ID. Short rest periods (fixation crosses of 2s) between sentences allowed the EEG to return to baseline and were not labeled. Over multiple runs (30 collected sessions of data), a dataset was built comprising numerous examples of EEG segments paired with the sentence that the user was reading or internally vocalizing. This procedure yielded a training corpus of a few thousand EEG-text pairs, with each EEG example being a 10-second, 16-channel signal segment associated with a known target sentence. Imagined speech decoding from EEG, while still a developing field, has seen growing interest and technical advances in recent years (Lopez-Bernal et al. [2022]).

To summarize sentence-locked responses and overall signal quality, we computed a grand-average event-related potential (ERP) over all sessions and electrodes (Figure 2). Signals were band-pass filtered (1–40 Hz), re-referenced to the average, and baseline corrected using the first 0.5 s of each epoch (processed data start at t = 0). The butterfly plot (all electrodes superimposed) reveals consistent, small-to-moderate deflections following sentence onset, while the scalp topographies at global field power (GFP) peaks provide a compact view of the evolving spatial pattern across the epoch. In total we averaged 3210 trials recorded with a 16-channel OpenBCI system.

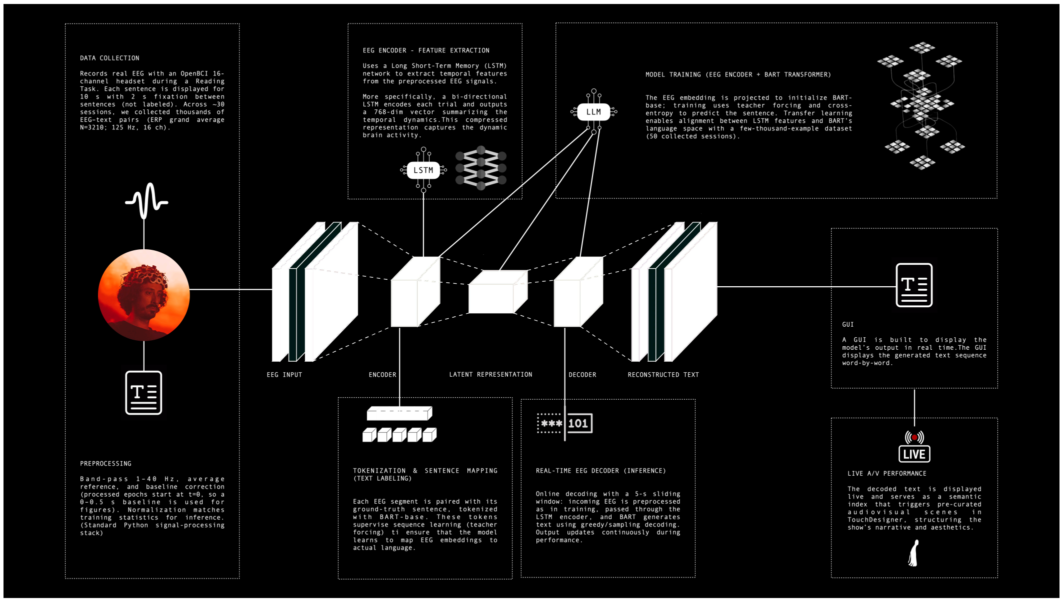

Figure 1: A cartography of *SYNAPTICON*.

To characterize induced (non–phase-locked) activity, we computed event-related spectral perturbations (ERSPs) from 2–40 Hz using Morlet wavelets (log-spaced frequencies; cycles scaled by frequency). For each trial we baseline-corrected power using the first 0.5 s of the epoch (processed data begin at t = 0) and expressed changes as log-ratios, then averaged across sessions and electrodes (channel for display selected in a centro-parietal region). The resulting time–frequency map shows a robust α-band (8–13 Hz) and β-band (13–30 Hz) desynchronization following sentence onset and persisting throughout reading, consistent with task engagement. Topographic insets at representative time–frequency points (α =10 Hz, β =21 Hz at 0.85 s) summarize the spatial distribution of these effects (Figure 3).

2.2 Model Architecture and Training:

The *SYNAPTICON* model, called EEGtoTextModel, consists of a bi-directional LSTM encoder that processes the EEG time-series and a pre-trained BART (the facebook/bart-base model) decoder that generates text. Concretely, the LSTM takes as input a sequence of EEG feature vectors (time points × 16 channels) and outputs a 768-dimensional hidden state vector at the final time-step. This 768-dim vector is used to initialize the BART decoder as if it were the first token's embedding (matching BART-base's hidden size of 768).

During training, the known sentence (target text) was fed into BART in teacher-forcing mode. That is, the model learned to predict the next word of the sentence given the EEG-derived context vector and the previous ground-truth words (Lu et al. [2024]). The training optimizes the sequence generation likelihood, effectively fine-tuning the BART model's ability to interpret the EEG encoder's representation. Importantly, BART's weights start from a model pre-trained on large text corpora, so it already has fluent language capability. We thus assume that only a relatively small number of epochs on the EEG dataset are needed to align the LSTM's neural features with BART's language space. Training on 30 sessions (a few thousand examples) is a small dataset for fine-tuning BART, but the model was still able to learn and steadily reduce its loss over epochs. This attests to the power of transfer learning: even limited EEG data can teach the model to associate certain brainwave patterns with words, because BART supplies the linguistic structure out-of-the-box.

2.3 Inference / Live Decoding:

After training, *SYNAPTICON* can operate in an online inference mode to convert incoming EEG signals into text in real time. The trained EEGtoTextModel is loaded and set to evaluation mode, and

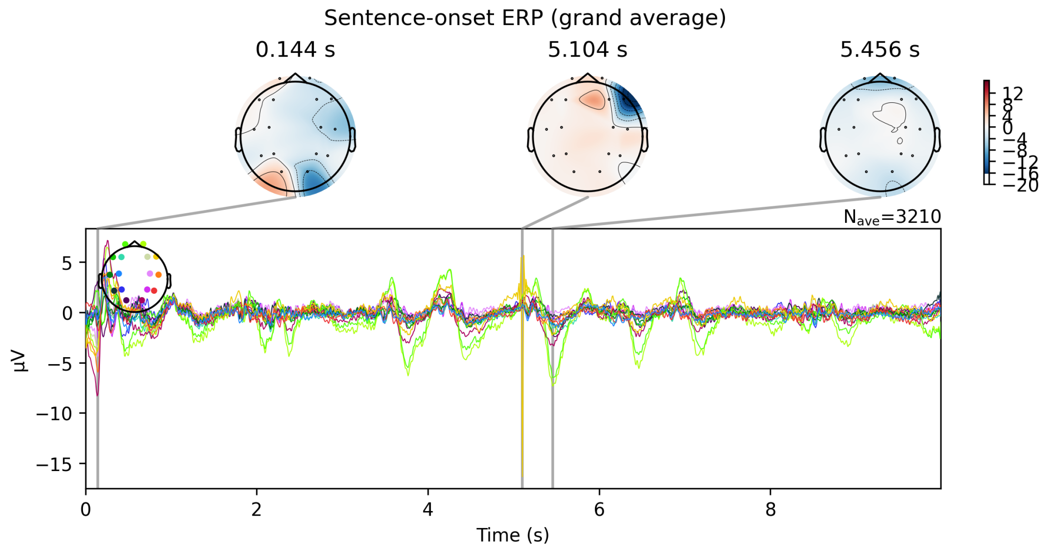

Figure 2: Sentence-onset ERP (grand average). All-channel butterfly plot (16 electrodes; N = 3210) and scalp topographies at data-driven GFP peaks (0.144 s, 5.104 s, 5.456 s). Data were band-pass filtered (1–40 Hz), average-referenced, and baseline-corrected using the first 0.5 s of each epoch (processed data begin at t = 0). The butterfly traces show the temporal profile of the sentence-locked response; topographies depict the corresponding spatial distributions at each peak latency.

the OpenBCI headset is connected to stream live EEG data. The system processes EEG in a sliding window of about 5 seconds (for instance, 625 samples at 125 Hz) – this window continuously moves, providing updated neural data chunks. Each 5-second EEG segment is preprocessed (band-pass filtered 1–40 Hz, normalized by the training set mean/std) exactly as the training data was. The segment is then fed through the LSTM encoder to produce a 768-dim vector, which is passed into the BART decoder to generate text output.

At inference, the BART model can generate words in a free-running manner, typically using greedy or sampling-based decoding to produce a sequence of words that it thinks corresponds to the EEG input. The system outputs the predicted text in a continuous fashion – for example, updating a graphical user interface (GUI) with new characters or words every few seconds as each window is processed. This real-time loop completes the BCI: the human simply thinks (reads or imagines) a sentence, and the system translates the brain activity into a textual output that appears on a screen, without any muscular intervention. Finally, this textual output, on top of being displayed per se, is also fed as a prompt into multimodal generative models to produce and orchestrate the corresponding show's audiovisual content. This entire project is open-source and all materials are available at the main author's GitHub repository, which contains a computationally lighter encoder for experimentation or deployment (faster training, fewer parameters).[1]

2.4 AI Multimodal Generation:

The output of the EEGtoTextModel functions as a semantic index for a pre-curated library of audiovisual "scenes", which triggers the whole artistic performance/live-show. The text output does not generate images in real time (e.g. via APIs), which is indeed part of our crucial next steps and further developments. In other words, when a new text segment is inferred from EEG data, the system uses it as a prompt for further analysis. Specific words are linked to the triggering of different scenes, so, depending on the live-decoded sentences, their activation determines the appearance of the subsequent scene and the evolution of the show. The pipeline was developed in TouchDesigner (by Derivative), a node-based visual programming language for real-time interactive multimedia content. This software not only enables the creation of a more adaptable and appropriately sized

[1] *SYNAPTICON*'s Open-Source GitHub repository: `https://github.com/AlbertBarqueDuran/SYNAPTICON-Brain-Waves-to-Natural-Language-to-Aesthetics`.

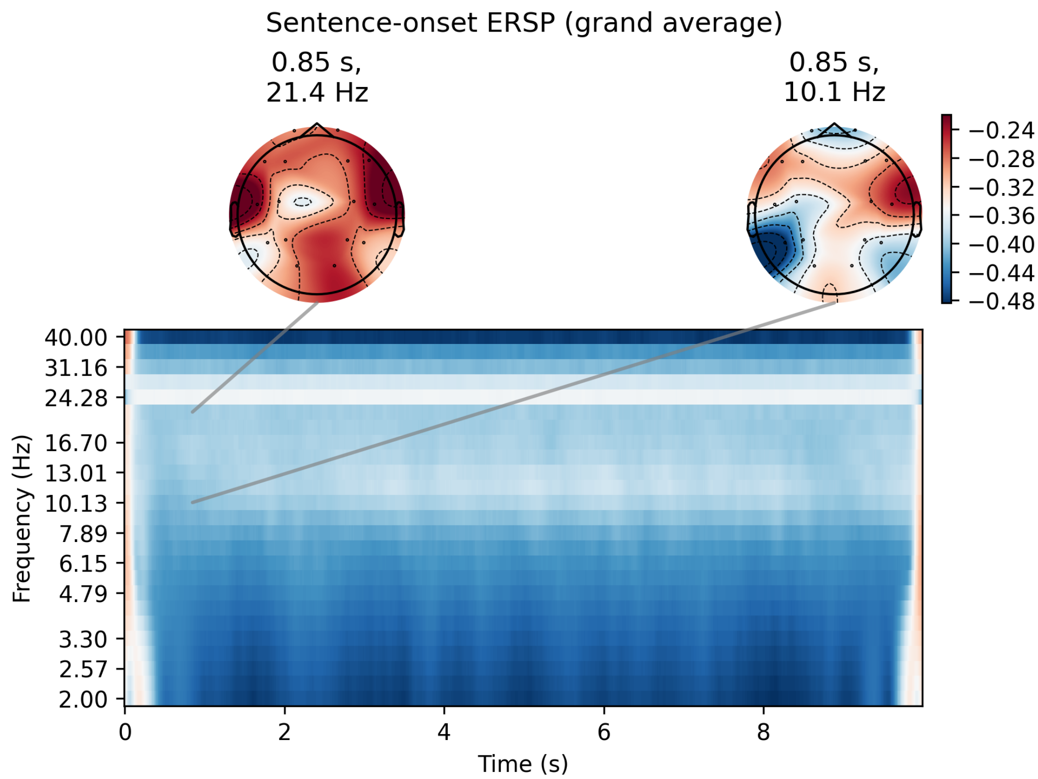

Figure 3: Sentence-onset ERSP (grand average). Time–frequency power (2–40 Hz) at a centro-parietal site, shown as log-ratio relative to a 0–0.5 s baseline. Cooler colors indicate power decreases (ERD) relative to baseline. A sustained $\alpha/beta$ desynchronization is visible during sentence processing. Top: scalp topographies at 0.85 s for β =21 Hz (left) and α =10 Hz (right) illustrate the spatial distribution of these band-limited effects.

project, but also allows for the development of shaders and patterns that respond to audio input. Each scene contains dedicated visual assets and shaders, designed and curated to align with the specific aesthetic and narrative of the show. This involves incorporating particular shapes, colors, textures, and visual effects to intentionally evoke various altered perceptual responses in the audience, such as hypnagogia, hallucinations and altered states of consciousness. In specific, Klüver Form Constants (i.e. tunnels, spirals, Honeycombs Gratings, Cobwebs) were used to produce the above-mentioned perceptual effects. This setup enables the performer's inner speech to structure not only the narrative sequence but also its sensorial expressivity, as every scene is prepared with different stroboscopic sequences and modulation of effects to create a unique experience for the audience.

Rather than aiming for a literal reconstruction of mental imagery, *SYNAPTICON* treats brain-derived text as a creative index, guiding a dynamic audiovisual environment that evolves in response to the performer's cognitive state. This establishes a closed loop of co-creation, in which internal thoughts continuously sculpt the shared artistic experience (Figure 4).

3 Discussion: Neuro-LLMs and its Case for Novel Human Expression and Alignment

3.1 Alignment at the Brain–Model Interfaces and other Neurotechnologies

SYNAPTICON addresses the "NeurIPS Creative AI Track 2025: HUMANITY" theme by exploring the profound symbiosis between biological and synthetic cognition. Foundation models can amplify human creativity, yet this symbiosis also exposes AI safety and alignment issues and critical concepts such as mental privacy. This necessitates the establishment of neurorights (i.e. cognitive liberty,

Figure 4: *SYNAPTICON* - Première Performance at Sónar Festival (Barcelona, Spain), June 2025.

mental integrity, and psychological continuity) as guardrails for signal-to-semantics technologies [Ienca and Andorno, 2017, Cassinadri, 2025].

When an LLM autocompletes a neural prompt, the extended mind thesis becomes tangible, as thought unfolds across both brain and model [Clark and Chalmers, 1998]. Designing with transparency, consent, and user agency can steer human–AI co-agency toward empowerment rather than exploitation. Ensuring brain–AI systems respect cognitive liberty is paramount—users must retain control over if and how their thoughts are decoded and used. *SYNAPTICON*'s pipeline highlights the need for interpretability and verifiability: users and observers should understand *what* the AI is inferring from EEG and *why*.

However, aligning AI behavior with human creative and ethical values is challenging because creativity and ethics itself are multifaceted [Barque-Duran and Pothos, 2021, Barque-Duran et al., 2018, 2016]. These considerations echo emerging AI safety work on the ethics of advanced AI assistants and principles that call for transparent, user-aligned systems and robust fail-safes [Amodei et al., 2016, Askell et al., 2021, Bostrom and Yudkowsky, 2014, Gabriel et al., 2014].

3.2 Open Problems and Next Steps

Looking ahead, we identify three fruitful directions: (i) *evaluation*—shared benchmarks for live EEG-to-text fidelity and audiovisual coherence; (ii) *governance*—practical neurorights-by-design patterns (on-device processing, permissioning, consent switches); and (iii) *co-creativity research*—studies on how human intent and model priors co-shape performance outcomes. By advancing along these lines, Neuro-LLM performance systems like *SYNAPTICON* can remain both visionary and responsible, extending human expression while safeguarding autonomy.

4 Acknowledgments

Authors want to thank all the partners that helped made possible the project: Sónar+D; OpenBCI; BSC (Barcelona Supercomputing Center); .NewArt foundation;; CBC (Center for Brain & Cognition) - Universitat Pompeu Fabra and the Departament de Cultura - Generalitat de Catalunya.

References

D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané. Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*, 2016.

A. Askell, Y. Bai, A. Chen, and et al. A general language assistant as a laboratory for alignment. *arXiv preprint arXiv:2112.00861*, 2021.

A. Barque-Duran and E. Pothos. Untangling decision routes in moral dilemmas: The refugees' dilemma. *American Journal of Psychology*, 2021. doi: 10.5406/amerjpsyc.134.2.0143.

A. Barque-Duran, E. Pothos, J. Yearsley, and J Hampton. Patterns and evolution of moral behavior: Moral dynamics in everyday life. *Thinking and Reasoning*, 2016. doi: 10.1080/13546783.2015.1051585.

A. Barque-Duran, E. Pothos, and J Yearsley, J.and Hampton. Contemporary morality: Moral judgments in digital contexts. *Computers in Human Behavior*, 2018. doi: 10.1016/j.chb.2017.05.020.

M Bostrom and E. Yudkowsky. *The ethics of artificial intelligence*. The Cambridge Handbook of Artificial Intelligence. Cambridge University Press, 2014.

G. Cassinadri. The right to mental integrity: Multidimensional, multilayered and extended. *Neuroethics*, 18(16), 2025. doi: 10.1007/s12152-025-09585-6.

A. Clark and D. J. Chalmers. The extended mind. *Analysis*, 58(1):7–19, 1998. doi: 10.1093/analys/58.1.7.

Y. Duan, J. Zhou, Z. Wang, Y. Wang, and C. Lin. Dewave: Discrete encoding of eeg waves for eeg to text translation. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

M. Ferrante, T. Boccato, F. Özçelik, R. VanRullen, and N. Toschi. Decoding the seeing brain: Reconstructing images and text from fmri recordings with multimodal diffusion models. In *NeurIPS Workshop on Diffusion Models*, December 2023.

I. Gabriel, A. Manzini, G. Keeling, L. A. Hendricks, V. Rieser, H. Iqbal, N. Tomašev, I. Ktena, Z. Kenton, M. Rodriguez, S. El-Sayed, S. Brown, C. Akbulut, A. Trask, E. Hughes, A. S. Bergman, R. Shelby, N. Marchal, and V... Bolina. The ethics of advanced ai assistants. *arXiv:2404.16244*, 2014.

H. Gürkök and A. Nijholt. Affective brain-computer interfaces for arts. In *Proc. ACII 2013*, pages 827–831, 2013.

M. Ienca and R. Andorno. Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, 13(5), 2017. doi: 10.1186/s40504-017-0050-1.

D. Lopez-Bernal, D. Balderas, P. Ponce, and A. Molina. A state-of-the-art review of eeg-based imagined speech decoding. *Frontiers in Human Neuroscience*, 16:867281, 2022. doi: 10.3389/fnhum.2022.867281.

J. Lu, Z. Yi, and X. Li. Enhancing eeg-to-text decoding through transferable representations from pretrained contrastive eeg-text masked autoencoders. In *Proceedings of ACL 2024*, 2024.

S. Murad, A. Dahal, and N. Rahimi. Eeg-to-text translation: A model for deciphering human brain activity. *arXiv preprint arXiv:2505.13936*, 2025.

A. Nijholt. *Brain Art: Brain-Computer Interfaces for Artistic Expression*. Springer, 2019.

F. Özçelik and R. VanRullen. Brain-diffuser: Natural scene reconstruction from fmri signals using generative latent diffusion. *Scientific Reports*, 13:15666, 2023. doi: 10.1038/s41598-023-42891-8.

L. Schreiner, A. Wipprecht, A. Olyanasab, S. Sieghartsleitner, H. Pretl, and C. Guger. Brain–computer-interface-driven artistic expression: real-time cognitive visualization in the pangolin scales animatronic dress and screen dress. *Frontiers in Human Neuroscience*, 19, 2025. doi: 10.3389/fnhum.2025.1516776.

J. Tang, M. Schrimpf, R. Eliahou, and A. J. Huth. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*, 26:1041–1050, 2023. doi: 10.1038/s41593-023-01304-9.

1 Supplemental Materials

1.1 Authors and Artists short biographies

Albert Barqué-Duran, PhD, is a cognitive scientist, AI researcher, and polymath exploring the future of human cognition, the Self, and ethics in Human-AI/AGI/ASI interactions. He is currently Research Scientist at the Interactive Arts and Science Laboratory (IASlab) at La Salle Barcelona (Universitat Ramon Llull), where he leads the PhD and MSc Programs in Creative Technologies. With over 15 peer-reviewed publications, his research combines computational cognitive science, experimental human–algorithm interaction, and critical AI studies to investigate machine intelligence, ethics, and the epistemological implications of human–AGI co-adaptation. He holds a PhD in Cognitive Science from the Centre for Mathematical Neuroscience at City St George's, University of London, and has held research appointments at the Moral Cognition Lab (Harvard University) and the Uehiro Centre for Practical Ethics (University of Oxford). Albert's work has been presented at leading cognitive science venues, AI–art symposia, festivals and biennales (e.g., CogSci, NeurIPS, Ars Electronica, Sónar+D, ISEA, MUTEK), and recognized by awards such as the Creative Europe Program, the Re:Humanism Prize, the Art of Neuroscience, and the LUMEN Prize Award (FUTURES), among others.

Ada M. Llauradó Crespo is a digital artist whose work explores the intersection of technology, cognition, and emotion. Starting from 3D experimentation, her practice has expanded to sculpture, jewelry, and interactive art, with a focus on extended cognition, neuroaesthetics, and real-time brain-computer interfaces (BCI). Her research investigates how identity, perception, and thought are fragmented and recombined through human–machine interaction. This is central to her collaborative master's thesis Multiplicity (with Ton Cortiella), where participants navigated an AI's latent space via EEG data. She was also part of the team behind Synapticon, a project led by Albert Barqué and presented at Sónar 2025, focused on EEG signal processing and artistic experimentation. She seeks to create poetic and dissonant spaces where science and art converge through experimental beauty.

Jesús Vaquerizo-Serrano artistically known as I AM JAS, is an electronic music artist and audio engineer currently based in Barcelona. As an acoustic engineer and researcher at Campus La Salle BCN (Ramon Llull University), he bridges technical expertise with a professional artistic career. His work has been featured in renowned venues such as Cova Santa Ibiza and Kappa FuturFestival, and released on acclaimed labels including Crosstown Rebels. His music has gained recognition from artists like MESTIZA and Black Coffee, and has been featured on official Spotify playlists such as Dunes.

Ton Cortiella Valls is a digital artist and creative technologist based in Barcelona, working at the intersection of generative systems, real-time visual programming, and creative AI. His practice focuses on immersive experiences and audiovisual interactivity, using tools such as TouchDesigner, Unreal Engine, Ableton Live, Blender, ComfyUI, and Max/MSP to build custom systems that blend procedural visuals with reactive sound. He has created visuals for international festivals like Atonal, Sónar, and MIRA, and contributed to commercial projects in cities like Las Vegas. With a background in Digital Arts and a Master's in Creative Technologies from La Salle Ramon Llull University, Ton combines computational design with emerging aesthetics. He is known for his collaborative spirit, curiosity, and ability to translate complex technical systems into emotional, sensory experiences.

Teo Ruffini Maiques is a multidisciplinary artist based in Barcelona working at the intersection of performance, dance, digital art, and music. A graduate in Digital Arts from La Salle – Universitat Ramon Llull, and with studies at LaSalle College of the Arts (Singapore) and La Massana (Barcelona), his work explores the dialogue between body and technology through hybrid formats. As a digital artist, he has exhibited works like ENAKD and *Spectral Resonator* at Expo Osaka 2025, Sónar Barcelona, BrainMind Summit (San Francisco), Colección SOLO (Madrid), and the Museu d'Art Contemporani de Barcelona. In performance, he has presented *TORERO* at IASLab and *To treat an object like it is a person...* at ONO (Singapore), among others. He has also performed in collaborative projects such as *Multiplicity* (with Ton Cortiella and Ada Llauradó) and *HUNTER* by Lluís Garau, and has danced with SVETLANA at venues including Paral·lel 62 in Barcelona.