AUDITING INDIVIDUAL SPECIFICITY IN EEG FOUNDATION MODELS WITH OCCUPANCY-PRESERVING TEMPORAL NULLS

Anonymous authors
Paper under double-blind review

ABSTRACT

EEG foundation models are typically evaluated through downstream predictive performance, providing limited insight into what individual-specific information their frozen representations preserve when transferred to an unseen dataset. In this study, we introduce a representation-audit framework that distinguishes *fingerprintability* from *temporal fingerprintability*. A temporal representation supports an individual-specific dynamical interpretation only if its cross-session fingerprintability exceeds an occupancy-preserving temporal null. We formalize this distinction through temporal excess, $\Delta_{\mathrm{temp}}(g) = \Delta_{\mathrm{FP}}^{\mathrm{original}}(g) - \mathbb{E}[\Delta_{\mathrm{FP}}^{\mathrm{permuted}}(g)]$. On the Temple University Epilepsy Corpus (TUEP), we evaluate two independently pretrained frozen EEG encoders, CNN–Transformer and LUNA-Base, and discretize their latent representations into $k{=}4$ states. State occupancy is reproducible across separate recording sessions ($N{=}112$), with $\Delta_{\mathrm{FP}}{=}0.103$ for CNN–Transformer and 0.096 for LUNA-Base. Empirical lag-1 transition representations W also appear individually specific under conventional metrics, reaching Top-1 retrieval of 5.8% and 4.5%, respectively, compared with 0.9% chance. In contrast, the apparent temporal fingerprintability does not exceed the occupancy-preserving temporal null: $\Delta_{\mathrm{temp}}(W) = -0.002$ for CNN–Transformer (original vs. permuted Δ_{FP}: 0.106 vs. 0.108) and -0.010 for LUNA-Base (0.095 vs. 0.105). Residual-transition and state-persistence probes yield the same conclusion at the primary $k{=}4$ resolution. These results show that a temporal representation can be individually fingerprintable without its fingerprintability depending on temporal order. More generally, $\Delta_{\mathrm{FP}} > 0$ does not imply $\Delta_{\mathrm{temp}} > 0$, motivating occupancy-preserving temporal nulls when interpreting individual-specific dynamics in learned representations.

1 INTRODUCTION

Electroencephalography (EEG) is a widely used noninvasive method for measuring brain activity and has been applied across neurological and clinical settings, including epilepsy assessment, sleep monitoring, and stroke research (Obeid & Picone, 2016; Jiang et al., 2024; Wang et al., 2025). Recent EEG foundation models (FMs) learn transferable latent representations from large-scale unlabeled corpora and are typically evaluated through frozen or lightly adapted downstream performance (Kostas et al., 2021; Jiang et al., 2024; Wang et al., 2025; Döner et al., 2026; Kuruppu et al., 2026; Zhou et al., 2026). This standard evaluation paradigm asks whether a pretrained representation extracts features that are useful for predicting predefined labels. However, clinical utility extends beyond population-level predictive performance. Human neurophysiology is inherently heterogeneous, and the heterogeneity can be particularly important when studying neurological disorders. For personalized and longitudinal applications, it is therefore important to understand not only whether a representation predicts a clinical label, but also whether it preserves reproducible individual-specific structure. Nevertheless, the extent to which individual-specific information survives the transformation from EEG signals into frozen FM latent spaces remains less understood. This creates an important gap between evaluating EEG FMs for downstream prediction and understanding their potential utility for individual-level clinical characterization. A second challenge is the interpretability of the latent space. Data-driven latent EEG states generally lack state-level clin-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ical ground truth (Kuruppu et al., 2026; Kastrati et al., 2025), making it difficult to determine what physiological or subject-specific information a discovered latent state actually represents. Consequently, separable latent states or strong downstream performance alone do not establish whether the learned representation captures stable individual characteristics, transient recording-specific variation, or clinically meaningful temporal structure. Pursuing predictive performance alone also does not reveal whether different encoders learn similar latent structures. Building on these limitations, we conduct a more direct examination of what individual-specific information is retained in frozen EEG FM representations. In particular, we ask whether an apparently dynamical signature reflects genuine temporal organization or simply stable state occupancy.

In this study, we develop a representation-audit framework using two lighter independently pretrained frozen EEG encoders: a CNN–Transformer-based encoder (Ahmadi & Mesin, 2025) and LUNA-Base (Döner et al., 2026). We use a *representation probe* in the diagnostic sense: a post-hoc analysis designed to characterize what information is preserved when a pretrained, frozen encoder is applied to an unseen clinical dataset, without updating the encoder or optimizing it for subject identity. For each encoder, a frozen mapping E_θ transforms EEG windows into latent embeddings, and unsupervised state discovery maps these embeddings to a discrete state sequence s_t. We then characterize each recording using complementary summaries of latent-state organization, including occupancy π, an empirical lag-1 transition matrix W, residual transitions $W^{\perp}$, mean dwell time, and self-transition probabilities P_{ii}.

Rather than asking only whether EEG FMs preserve "individual dynamics," we ask two more fundamental questions:

1. **Do frozen EEG representations preserve individual specificity?**
2. **Which components carry this information?**

We distinguish two levels of information in the resulting latent-state sequences. State occupancy π describes *which states* an individual tends to occupy and how frequently they occur, whereas temporal summaries describe *how those states are ordered over time*. Persistence measures such as dwell time and self-transition probability provide complementary diagnostics of this distinction. Notably, these components are not independent. A transition-based representation W can depend on both state occupancy and temporal ordering:

$$W = g(\pi, \mathcal{T}),$$

where π denotes state occupancy and $\mathcal{T}$ denotes temporal organization. Thus, individual specificity in W may arise from

$$\underbrace{\pi}_{\text{state occupancy}} \quad \text{or} \quad \underbrace{\mathcal{T}}_{\text{temporal organization}} .$$

We therefore distinguish general *fingerprintability*, which asks whether a representation contains reproducible individual information, from *temporal fingerprintability*, which asks whether temporal order contributes individual information beyond occupancy.

Contributions.

- We introduce a representation audit that separates general individual specificity from individual information attributable to temporal ordering beyond state occupancy, using an occupancy-preserving temporal null.
- Across two independently pretrained frozen EEG encoders, we find reproducible individual information across separate recording sessions in the Temple University Hospital EEG Epilepsy Corpus (TUEP), a dataset distinct from the Temple University Hospital EEG Corpus (TUEG) pretraining corpus. This shows that subject-specific patterns of latent-state occupancy can be preserved during cross-dataset transfer without subject-identity supervision.
- We show that individual specificity in transition representations does not necessarily reflect temporal organization: it can persist even after temporal order is destroyed. This demonstrates that apparent dynamical fingerprints can be driven by stable state occupancy, motivating occupancy-preserving temporal nulls for evaluating individual-specific dynamics.

2 Related Work

Our work connects EEG foundation models and representation analysis, individual specificity in EEG, and latent-state dynamics.

EEG foundation models and representation analysis. Recent self-supervised EEG encoders include BENDR (Kostas et al., 2021), BIOT (Yang et al., 2023), Brant (Zhang et al., 2023), LaBraM (Jiang et al., 2024), CBraMod (Wang et al., 2025), Neuro-GPT (Cui et al., 2024), REVE (El Ouahidi et al., 2026), and LUNA (Döner et al., 2026). These models are commonly evaluated through downstream prediction and transfer performance, while recent work has increasingly examined the structure and interpretability of their learned representations (Kuruppu et al., 2026; Kastrati et al., 2025). Our work follows this direction but asks a different question: what individual-specific information is retained when pretrained EEG representations are frozen and transferred to a separate clinical dataset?

Individual specificity in EEG. EEG has long been known to contain subject-specific information, motivating work on person recognition and EEG biometrics (Marcel & Millán, 2007; DelPozo-Banos et al., 2015; Maiorana, 2020). Prior studies have further examined the stability of such information across recording sessions, highlighting both its reproducibility and sensitivity to session-related variability (DelPozo-Banos et al., 2015; de Ataide et al., 2026). At the same time, EEG captures neurophysiological alterations associated with brain disorders, including epilepsy, stroke, and neurodegenerative diseases (Tabbal et al., 2025; Williamson et al., 2024). More recently, subject-aware EEG representation learning has explicitly used subject identity as a supervisory signal, including contrastive objectives that treat recordings from the same subject as positive pairs (Karantonis et al., 2025). Unlike these approaches, we do not optimize the encoder or a classifier using subject identity. Instead, we use cross-session similarity as a post-hoc probe of individual information already present in representations produced by frozen encoders. We therefore study *individual specificity* rather than biometric identification.

Occupancy and temporal dynamics. Latent-state models are widely used to summarize time-varying neural activity, with state occupancy characterizing how frequently a state is visited and transition statistics characterizing how states evolve over time (Vidaurre et al., 2017; Quinn et al., 2018). Those summaries have been used to characterize temporal organization in both EEG and other neuroimaging modalities. However, occupancy and empirical transition statistics are not independent: a transition representation can retain information about stable state use even after the original temporal ordering is removed. Consequently, cross-session reproducibility of a transition representation alone does not establish individual-specific temporal organization. We address this ambiguity using an occupancy-preserving temporal null that keeps state counts fixed while destroying temporal order.

3 Background

We audit two independently pretrained frozen EEG encoders: a CNN–Transformer (1.46M parameters, 128-d) adapted from Ahmadi & Mesin (2025), and LUNA-Base (Döner et al., 2026) (7.10M, 256-d). Neither encoder is fine-tuned on TUEP or trained with subject identity. Architecture, preprocessing, pretraining, and frozen transfer performance are in Appendix A.

Each 30-s window x_t is mapped to $z_t = E_\theta^{(e)}(x_t)$. Global k-means (MacQueen, 1967) assigns a discrete state $s_t \in \{1, \dots, k\}$ independently in each latent space(k=4 primary; $k \in \{3, 5, 6\}$ in Appendix G). Occupancy of state i in a session of length T is

$$\pi_i = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}[s_t = i]. \tag{1}$$

Thus π records **which** states are used and **how often**, not their order.

The empirical lag-1 transition matrix is column-stochastic,

$$\widehat{P}_{ji} = \frac{\sum_{t=2}^{T} \mathbf{1}[s_{t-1} = i,\ s_t = j]}{\sum_{t=2}^{T} \mathbf{1}[s_{t-1} = i]}, \tag{2}$$

so $\widehat{P}_{ji} = P(s_t = j \mid s_{t-1} = i)$. We vectorize it as $W = \mathrm{vec}(\widehat{P}) \in \mathbb{R}^{k^2}$ (column-major; $W \in \mathbb{R}^{16}$ at k=4). The explicit stacking and residual/dwell formulas are in Appendix B. Because $\widehat{P}$ is estimated from states with subject-specific occupancy, a positive $\Delta_{\mathrm{FP}}(W)$ need not mean that temporal order is individual-specific. That distinction is formalized next.

4 Methods and Problem formulation

4.1 Cross-session individual specificity (Δ_{FP})

Our central question is whether a frozen EEG representation preserves **temporal individual specificity**: individual-specific information that depends on the temporal ordering of latent states rather than only on how frequently those states occur.

For the primary analysis, each subject s contributes two separate recording sessions, denoted A_s and B_s. Let v_s^A and v_s^B denote vector signatures derived from these sessions, where v may represent either state occupancy π or a temporal representation such as $W = \mathrm{vec}(\widehat{P})$.

We quantify similarity between two signatures u and v using cosine similarity,

$$S(u, v) = \frac{u^\top v}{\|u\|_2 \|v\|_2}. \tag{3}$$

The within-subject and between-subject similarities are defined as

$$S_{\mathrm{within}}(s) = S(v_s^A, v_s^B), \tag{4}$$

and

$$S_{\mathrm{between}}(s, r) = S(v_s^A, v_r^B), \qquad r \neq s, \tag{5}$$

respectively.

We quantify individual specificity using the fingerprintability gap

$$\Delta_{\mathrm{FP}} = \mathbb{E}_s \left[S_{\mathrm{within}}(s) \right] - \mathbb{E}_{s \neq r} \left[S_{\mathrm{between}}(s, r) \right]. \tag{6}$$

A positive Δ_{FP} indicates that representations from separate sessions of the same subject are, on average, more similar than representations from different subjects. Thus, $\Delta_{\mathrm{FP}} > 0$ indicates reproducible individual information, but does not establish that this information depends on temporal ordering.

As a complementary measure of individual specificity, we perform cross-session subject retrieval. For each subject s, the session-A signature v_s^A is used as a query and compared by cosine similarity against the session-B signatures of all N subjects, $\{v_r^B\}_{r=1}^N$. Top-1 retrieval is correct when the corresponding same-subject signature v_s^B is ranked first. We repeat the procedure in the reverse direction, querying each session-B signature against all session-A signatures, and average the retrieval rates across the two directions. For each query, exactly one of the N candidate signatures corresponds to the same subject. Therefore, random selection would identify the correct subject with probability $1/N$ at Top-1.

4.2 Occupancy-preserving temporal null (Δ_{temp})

From the previous section, Δ_{FP} provides a measure of individual specificity; however, a positive Δ_{FP} does not necessarily imply that temporal ordering itself is individual-specific. In particular, transition statistics can inherit subject-specific information from state occupancy. Therefore, we construct an occupancy-preserving temporal null to separate individual specificity due to state use from that due to temporal ordering.

For each recording, we independently permute the latent-state sequence while preserving its state counts. For example,

$$\begin{aligned} \text{Original:} \quad & 1,\ 1,\ 2,\ 1,\ 3,\ 3,\ 2,\ 1, \\ \text{Permuted:} \quad & 3,\ 1,\ 2,\ 1,\ 1,\ 3,\ 1,\ 2. \end{aligned}$$

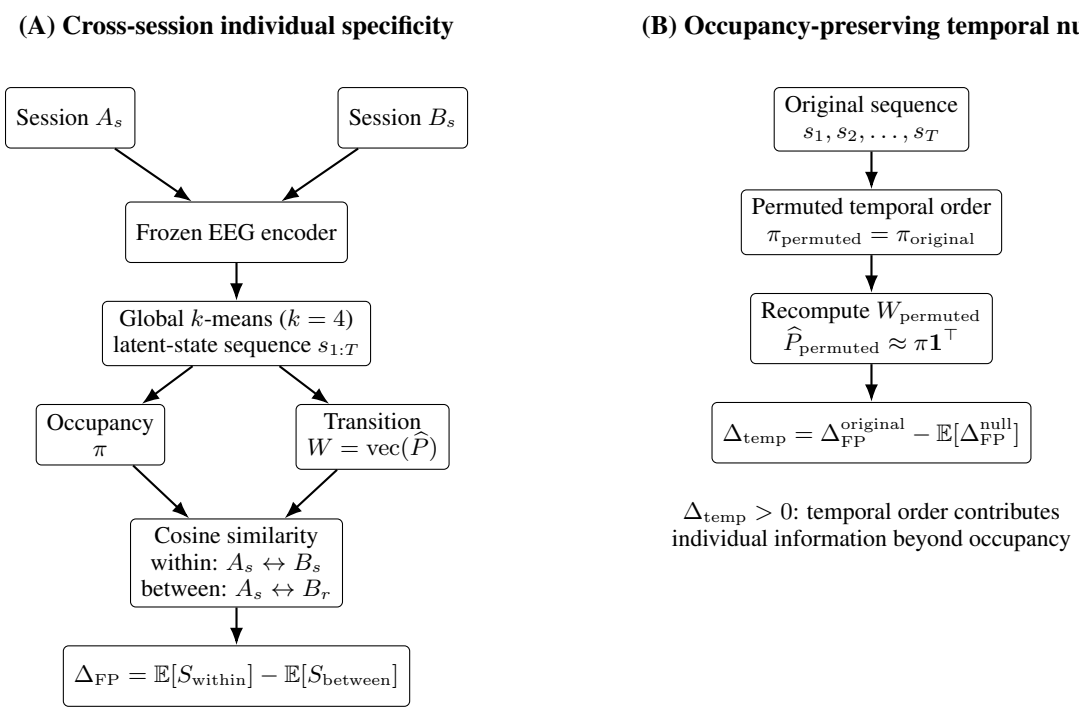


Figure 1: **Representation-audit framework.** (A) Separate recording sessions from the same subject are encoded using a frozen EEG encoder and mapped to $k = 4$ latent states. Occupancy π and the lag-1 transition representation W are compared across sessions to measure individual specificity using Δ_{FP}. (B) The occupancy-preserving temporal null keeps state counts fixed while destroying temporal order. Δ_{temp} measures whether individual specificity exceeds that expected from occupancy alone.

The two sequences contain the same number of occurrences of each state and therefore have identical occupancy, while their temporal ordering differs.

$$\pi_{\mathrm{original}} = \pi_{\mathrm{permuted}}, \tag{7}$$

For a candidate temporal signature g, we define temporal excess as the difference between its fingerprintability in the original sequences and its average fingerprintability under occupancy-preserving temporal shuffling:

$$\Delta_{\mathrm{temp}}(g) = \Delta_{\mathrm{FP}}^{\mathrm{original}}(g) - \mathbb{E}\left[\Delta_{\mathrm{FP}}^{\mathrm{permuted}}(g)\right]. \tag{8}$$

While $\Delta_{\mathrm{FP}} > 0$ quantifies reproducible individual information, $\Delta_{\mathrm{temp}} > 0$ indicates that temporal ordering contributes individual-specific information beyond that retained by occupancy alone. Occupancy itself provides a sanity check because permutation leaves π unchanged, giving $\Delta_{\mathrm{temp}}(\pi) = 0$ by construction. For a temporal signature such as W, shuffling changes the transition structure while preserving state occupancy. We therefore quantify the contribution of the original temporal ordering as

$$\Delta_{\mathrm{temp}}(W) = \Delta_{\mathrm{FP}}(W_{\mathrm{original}}) - \mathbb{E}\left[\Delta_{\mathrm{FP}}(W_{\mathrm{permuted}})\right].$$

A positive $\Delta_{\mathrm{temp}}(W)$ indicates that the original temporal ordering contributes individual-specific information beyond occupancy. Figure 1 summarizes the representation-audit framework for Δ_{FP} and Δ_{temp}.

4.3 Temporal signatures and tests

The primary temporal object is $W = \mathrm{vec}(\widehat{P})$. Transitions are counted only within a recording. After occupancy-preserving shuffling, the next-state distribution approaches occupancy, $\widehat{P}(\cdot \mid i) \approx \pi$, so

$$\widehat{P}_{\mathrm{permuted}} \approx \pi \mathbf{1}^\top. \tag{9}$$

Thus W_{permuted} can remain individually specific through π even though the matrices differ. $\Delta_{\text{temp}}(W)$ tests whether fingerprintability depends on the original order beyond occupancy.

Secondary probes, residual transitions $W^{\perp} = \text{vec}(\widehat{P} - \pi\mathbf{1}^{\top})$, mean dwell, $\text{dwell}^{\perp}$, and P_{ii}, are defined in Appendix B and tested in Appendix D.

We report Δ_{FP}, Top-1, and a two-sided Mann–Whitney test on within- versus between-subject similarity. The temporal test is one-sided:

$$H_0 : \Delta_{\text{temp}}(g) \leq 0 \qquad \text{vs.} \qquad H_1 : \Delta_{\text{temp}}(g) > 0.$$

Using 100 occupancy-preserving permutations, we compute $p_{\text{perm}} = \Pr(\Delta_{\text{FP}}(g_{\text{permuted}}) \geq \Delta_{\text{FP}}(g_{\text{original}}))$. A large p_{perm} indicates that comparable fingerprintability can be obtained without the original temporal order. Robustness checks (block shuffle, split-half, duration, k, and cross-encoder analyses) are provided in the Appendix.

5 Experimental setup

We use TUEP (Obeid & Picone, 2016) with n=198 labeled subjects (99 epilepsy / 99 non-epilepsy). Preprocessing (Appendix A) yields 2,571 recordings and 73,884 30-s windows ($\approx$ 616 hours). Each 30-s window is passed through a frozen encoder to obtain a latent embedding, and a shared global k-means model assigns each embedding to one of k latent states. The temporally ordered state assignments within each recording define the latent-state sequence $s_{1:T}$. The primary analysis uses N=112 subjects (75 epilepsy / 37 non-epilepsy) with at least two sessions of $\geq$ 20 windows; we take the two longest sessions A_s and B_s. Within-subject pairs are $A_s \leftrightarrow B_s$; between-subject pairs are $A_s \leftrightarrow B_r$ for $r \neq s$. Subject identity is used only for this evaluation. All primary numbers use frozen CNN–Transformer or LUNA-Base, global k=4, cosine similarity, and these N=112 pairs. Chance Top-1 is $1/112 \approx 0.0089$.

6 Results

The results answer the two questions in Section 1 in order. We first ask whether any individual-specific information is preserved across sessions. We then ask which component of the representation carries that information: occupancy, persistence, or temporal order beyond occupancy.

6.1 Occupancy is individually specific

We first ask whether the simplest property of the latent-state representation, state occupancy, preserves reproducible individual information across recording sessions. Occupancy is more similar within subjects than between subjects for both encoders (Table 1). CNN–Transformer yields $\Delta_{\text{FP}}(\pi)$=0.103 (p=3.9 $\times$ 10^{-4}; Top-1 = 0.018), while LUNA-Base yields $\Delta_{\text{FP}}(\pi)$=0.096 (p=8.3 $\times$ 10^{-4}; Top-1 = 0.009).

These results indicate modest but statistically detectable cross-session individual specificity in the occupancy signature π: subjects tend to occupy similar mixtures of latent states across separate recording sessions. However, Top-1 retrieval remains low (1.8% for CNN–Transformer and 0.9% for LUNA-Base, compared with 0.9% chance). We therefore interpret π as evidence of individual-specific information in representations produced by the frozen encoders rather than as a biometric identifier. Importantly, the occupancy-preserving temporal null retains each recording's state counts by construction, such that $\pi_{\text{permuted}} = \pi_{\text{original}}$. Thus, the individual-specific state-use information observed in Table 1 is explicitly preserved under the null. The subsequent comparison of original and permuted transition signatures $W = \text{vec}(\widehat{P})$ therefore tests whether temporal ordering contributes additional individual information beyond that retained under the occupancy-preserving control.

6.2 A transition representation appears fingerprintable

We next examine whether the empirical lag-1 transition representation $W = \text{vec}(\widehat{P})$ is reproducible across sessions. Like occupancy, W exhibits a positive cross-session fingerprintability gap. CNN–Transformer yields Δ_{FP}=0.106 with Top-1 = 0.058, while LUNA-Base yields Δ_{FP}=0.095 with

Sig.	Enc.	Within / between	Δ_{FP}	Top-1	MW p	Δ_{temp}
π	CNN–Transformer	0.586/0.483	0.103	0.018	3.9×10^{-4}	0
π	LUNA-Base	0.594/0.498	0.096	0.009	8.3×10^{-4}	0
W	CNN–Transformer	0.554/0.448	0.106	0.058	6.4×10^{-5}	−0.002
W	LUNA-Base	0.547/0.453	0.095	0.045	9.9×10^{-5}	−0.010
	Enc.	Original Δ_{FP}	Permuted Δ_{FP}	Permuted Top-1	p_{perm}	
W permute	CNN–Transformer	0.106	0.108±0.008	0.042	0.63	
W permute	LUNA-Base	0.095	0.105±0.006	0.038	0.92	

Table 1: Cross-session individual-specificity audit on TUEP (k=4, N=112; chance Top-1 = 0.009). Permute denotes 100 occupancy-preserving temporal shuffles. Δ_{temp} is the difference between the original Δ_{FP} and the mean permuted Δ_{FP}. For occupancy, $\Delta_{\mathrm{temp}}(\pi) = 0$ by construction. For W, Δ_{temp} is estimated from 100 shuffles and is not forced to zero. Occupancy provides a positive individual-specific signal, whereas the empirical lag-1 transition representation W does not exceed its temporal null.

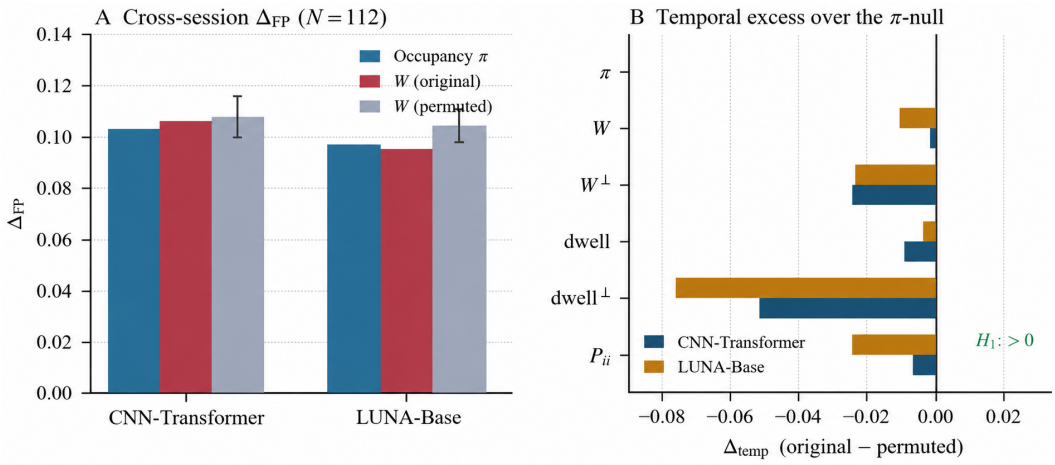


Figure 2: Cross-session individual-specificity audit on TUEP (k=4, N=112). (A) Occupancy Δ_{FP} is comparable to that of the empirical lag-1 transition representation W, while an occupancy-preserving permutation of W matches or exceeds the original W gap (error bars: standard deviation over 100 shuffles). (B) At the primary k=4 setting, none of the examined temporal signatures shows positive excess over the occupancy-preserving temporal null. Occupancy provides the positive individual-specific baseline, while $\Delta_{\mathrm{temp}}(\pi) = 0$ by construction.

Top-1 = 0.045 (Table 1). For CNN–Transformer, Top-1 retrieval is approximately six times the chance level of 0.0089.

Considered alone, these within- versus between-subject results could suggest that the latent transition structure carries a reproducible individual signature. However, $\Delta_{\mathrm{FP}}(W)$ is similar in magnitude to the occupancy fingerprintability gap. This observation motivates a stricter question: does the individual specificity of W actually depend on the original temporal ordering of the latent states, or can it arise from subject-specific state occupancy retained by the transition matrix?

6.3 Transition Fingerprintability under the Null

We address this question using the temporal null that preserves occupancy as defined in Section 4.2. Within each recording, state labels are randomly permuted while preserving the number of occurrences of every state. The permuted sequence therefore has exactly the same occupancy as the original sequence but no longer preserves its original temporal ordering. The empirical lag-1 transition matrix and W are then recomputed from each permuted sequence. Table 1 and Figure 2 show the results of testing whether the fingerprintability of the transition sig-

nature W depends on the original temporal ordering. Under H_1, the original sequence should yield greater fingerprintability than the occupancy-preserving permuted sequences, corresponding to $\Delta_{\mathrm{temp}}(W) > 0$. However, we do not observe this pattern at the primary $k{=}4$ resolution. For CNN–Transformer, $\Delta_{\mathrm{FP}}(W_{\mathrm{original}}) = 0.106$, whereas the occupancy-preserving shuffles yield $\mathbb{E}[\Delta_{\mathrm{FP}}(W_{\mathrm{permuted}})] = 0.108 \pm 0.008$ ($p_{\mathrm{perm}} = 0.63$), giving $\Delta_{\mathrm{temp}}(W) = -0.002$. For LUNA-Base, the corresponding values are 0.095 and 0.105 ± 0.006 ($p_{\mathrm{perm}} = 0.92$), giving $\Delta_{\mathrm{temp}}(W) = -0.010$. Thus, neither encoder provides evidence for H_1 at $k{=}4$.

Importantly, shuffling changes the transition matrix by destroying the original temporal ordering, such that $W_{\mathrm{original}} \neq W_{\mathrm{permuted}}$ in general. But it preserves each recording's state occupancy, $\pi_{\mathrm{permuted}} = \pi_{\mathrm{original}}$. Under the permuted sequence,

$$\widehat{P}_{\mathrm{permuted}} \approx \pi \mathbf{1}^{\top},$$

so the transition representation can still retain information about which latent states a subject tends to occupy. Therefore, although the transition matrices change after shuffling, their cross-session fingerprintability remains comparable.

The key distinction is therefore between fingerprintability and temporal fingerprintability:

$$\Delta_{\mathrm{FP}}(W) > 0 \not\Rightarrow \Delta_{\mathrm{temp}}(W) > 0 \tag{10}$$

A positive $\Delta_{\mathrm{FP}}(W)$ indicates that the transition representation contains reproducible individual information. In contrast, a positive $\Delta_{\mathrm{temp}}(W)$ would indicate that the original temporal ordering contributes individual information beyond that retained under the occupancy-preserving null. Our primary results support the former but not the latter.

This conclusion is not specific to W. At $k{=}4$, residualized and persistence-based signatures, including $W^{\perp}$, dwell, $\mathrm{dwell}^{\perp}$, and P_{ii}, likewise do not show positive temporal excess (Appendix D). Additional split-half, long-recording, state-resolution, and cross-encoder analyses are reported in the Appendix. Together, these analyses support the primary conclusion that the tested temporal signatures do not provide evidence for individual-specific temporal ordering beyond the occupancy-preserving null.

7 Discussion

This study separates two questions that are often conflated when analyzing learned temporal representations: whether a frozen representation preserves reproducible individual information, and whether that information depends specifically on temporal order. Across two independently pretrained EEG encoders, latent-state occupancy shows modest but reproducible individual specificity across separate recording sessions. Transition representations W also appear individually specific under conventional fingerprinting metrics; however, their fingerprintability does not exceed the occupancy-preserving temporal null at the primary $k{=}4$ resolution. Specifically, $\Delta_{\mathrm{temp}}(W) = -0.002$ for CNN–Transformer and -0.010 for LUNA-Base. Thus, a temporal representation can be individually fingerprintable without its fingerprintability depending on the original temporal order.

This distinction can be understood from the relationship between occupancy and empirical transition statistics. Because $W = g(\pi, \mathcal{T})$, where π denotes state occupancy and $\mathcal{T}$ denotes temporal ordering, individually specific occupancy can propagate into W. After temporal shuffling, $\widehat{P}_{\mathrm{permuted}} \approx \pi \mathbf{1}^{\top}$, such that the transition representation can continue to encode which latent states a subject tends to occupy even after the original temporal order has been destroyed. Consequently, positive $\Delta_{\mathrm{FP}}(W)$, significant within-versus-between similarity differences, or above-chance retrieval are not sufficient by themselves to establish individual-specific temporal organization. The temporal-excess test in Equation 8 directly addresses this ambiguity by asking whether fingerprintability exceeds that obtained under an occupancy-preserving temporal null.

The primary findings therefore support modest individual specificity in latent-state occupancy, but do not provide evidence that lag-1 W, $W^{\perp}$, or $\mathrm{dwell}^{\perp}$ captures individual-specific temporal ordering beyond the null at $k{=}4$. These findings should also not be interpreted as evidence for biometric identification or as establishing W as a causal network. Instead, they motivate occupancy-preserving temporal nulls as a diagnostic tool for distinguishing individual-specific state use from individual-specific temporal organization. Importantly, this principle is not specific to EEG and may be useful when interpreting other learned sequential representations.

The absence of temporal excess does not imply that the latent-state sequences lack clinically relevant structure. As shown in Appendix H, epilepsy and non-epilepsy groups differ in occupancy entropy and mean self-transition probability, whereas $\|W^{\perp}\|_F$ does not differ significantly between groups. Thus, population-level clinical sensitivity and subject-specific temporal organization represent distinct properties of the learned representation.

However, several limitations should be considered. First, the primary analysis is based on a single clinical EEG corpus, and replication across additional datasets and recording conditions is needed. Second, many recordings are relatively short, which may increase variability in empirical transition and dwell estimates, although the long-recording analysis provides an additional sensitivity check. Third, absolute Top-1 retrieval remains modest, limiting any interpretation in terms of biometric identification. Fourth, the study examines two frozen encoders and uses a discrete $k{=}4$ latent-state representation as the primary probe. Sensitivity analyses across $k \in \{3, 5, 6\}$ broadly support the primary conclusion (Appendix G), although LUNA-Base at $k{=}6$ shows a small positive temporal excess that is not replicated by CNN–Transformer. Therefore, the present findings should be interpreted as a diagnostic of the representations examined here rather than as a universal claim about individual-specific EEG dynamics.

8 Conclusion

We introduced a representation audit that separates general individual specificity from individual information attributable to temporal ordering beyond state occupancy. Across two frozen EEG encoders, latent-state occupancy preserves modest but reproducible individual information across separate recording sessions. Although empirical transition representations also appear individually fingerprintable, their fingerprintability does not exceed an occupancy-preserving temporal null at the primary $k{=}4$ resolution. Thus, fingerprintability of a temporal representation should not by itself be interpreted as evidence of individual-specific temporal organization. More broadly, our results motivate occupancy-preserving temporal nulls as a complementary diagnostic for learned sequential representations:

$$\Delta_{\mathrm{FP}}(W) > 0 \;\not\Rightarrow\; \Delta_{\mathrm{temp}}(W) > 0.$$

Acknowledgments

Hidden for double-blind review.

AI Use Statement

AI tools were used to polish writing, provide suggestions on manuscript organization, and assist with debugging portions of the analysis code. All numerical results were computed from the authors' state-sequence exports and verified by the authors. AI was not used to generate or fabricate experimental results.

Ethics statement

Analyses use the publicly released TUEP clinical EEG corpus. No new human-subjects enrollment was performed. We do not present the method as a biometric identification system.

Reproducibility statement

Preprocessor settings, encoder sizes, $k{=}4$ as primary, signature definitions, Δ_{temp}, and occupancy-preserving nulls are specified in Sections 3–5 and Appendices A–B. Occupancy, lag-1 W, residual, dwell, split-half, and temporal-null tests are produced from per-recording $k{=}4$ sequences (one `.npz` per subject). Sensitivity $k \in \{3, 5, 6\}$ uses the same layout. Secondary group comparisons pool labeled recordings within subject and do not enter the primary Δ_{temp} test. Analysis scripts and the per-subject state sequences will be released upon acceptance.

REFERENCES

Hossein Ahmadi and Luca Mesin. Universal semantic feature extraction from eeg signals: a task-independent framework. *Journal of Neural Engineering*, 22(3):036003, 2025.

Wenhui Cui, Woojae Jeong, Philipp Thölke, Takfarinas Medani, Karim Jerbi, Anand A Joshi, and Richard M Leahy. Neuro-gpt: Towards a foundation model for eeg. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5. IEEE, 2024.

Marissa L de Ataide, Narayan Vetrekar, Krishna Patel, Rajendra Gad, and Raghavendra Ramachandra. Brain signal for secure eeg biometric authentication: A comprehensive survey. *Sensors (Basel, Switzerland)*, 26(13):4045, 2026.

Marcos DelPozo-Banos, Carlos M Travieso, Christoph T Weidemann, and Jesús B Alonso. Eeg biometric identification: a thorough exploration of the time-frequency domain. *Journal of neural engineering*, 12(5):056019, 2015.

Berkay Döner, Thorir Mar Ingolfsson, Luca Benini, and Yawei Li. LUNA: efficient and topology-agnostic foundation model for EEG signal analysis. *Advances in Neural Information Processing Systems*, 38:70682–70708, 2026.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. Reve: A foundation model for eeg-adapting to any setup with large-scale pretraining on 25,000 subjects. *Advances in Neural Information Processing Systems*, 38:22541–22577, 2026.

Wei-Bang Jiang, Liming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous EEG data in BCI. In *International Conference on Learning Representations*, 2024.

Antonis Karantonis, Konstantinos Barmpas, Dimitrios Adamos, Nikolaos Laskaris, Stefanos Zafeiriou, and Yannis Panagakis. Subject-aware contrastive learning for eeg foundation models. In *NeurIPS 2025 Workshop on Learning from Time Series for Health*, 2025.

Ard Kastrati, Josua Bürki, Jonas Lauer, Cheng Xuan, Raffaele Iaquinto, and Roger Wattenhofer. Eeg-bench: A benchmark for eeg foundation models in clinical applications. *arXiv preprint arXiv:2512.08959*, 2025.

Demetres Kostas, Stephane Aroca-Ouellette, and Frank Rudzicz. Bendr: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of eeg data. *Frontiers in Human Neuroscience*, 15:653659, 2021.

Gayal Kuruppu, Neeraj Wagh, Vaclav Kremen, and Yogatheesan Varatharajah. Eeg foundation models: a critical review of current progress and future directions. *Journal of neural engineering*, 23(2):021001, 2026.

James MacQueen. Some methods for classification and analysis of multivariate observations. 1: 281–297, 1967.

Emanuele Maiorana. Deep learning for EEG-based biometric recognition. *Neurocomputing*, 410: 374–386, 2020.

Sebastien Marcel and José del R Millán. Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 29(4):743–752, 2007.

Iyad Obeid and Joseph Picone. The Temple University Hospital EEG data corpus. *Frontiers in Neuroscience*, 10:196, 2016.

Andrew J Quinn, Diego Vidaurre, Romesh Abeysuriya, Robert Becker, Anna C Nobre, and Mark W Woolrich. Task-evoked dynamic network analysis through hidden markov modeling. *Frontiers in neuroscience*, 12:603, 2018.

Judie Tabbal, Aida Ebadi, Ahmad Mheich, Aya Kabbara, Bahar Güntekin, Görsev Yener, Veronique Paban, Ute Gschwandtner, Peter Fuhr, Marc Verin, et al. Characterizing the heterogeneity of neurodegenerative diseases through eeg normative modeling. *npj Parkinson's Disease*, 11(1): 117, 2025.

Diego Vidaurre, Stephen M Smith, and Mark W Woolrich. Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, 114(48):12827–12832, 2017.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. CBraMod: A criss-cross brain foundation model for EEG decoding. In *International Conference on Learning Representations*, 2025.

Jordan N Williamson, Beni E Mulyana, Rita Huan-Ting Peng, Dorothy He, Shirley A James, Evgeny V Sidorov, and Yuan Yang. Targeted brain stimulation alters resting state eeg and reduces post-stroke impairment. In *International Conference on NeuroRehabilitation*, pp. 357–361. Springer, 2024.

Chaoqi Yang, M Westover, and Jimeng Sun. Biot: Biosignal transformer for cross-data learning in the wild. *Advances in Neural Information Processing Systems*, 36:78240–78260, 2023.

Daoze Zhang, Zhizhang Yuan, Yang Yang, Junru Chen, Jingjing Wang, and Yafeng Li. Brant: Foundation model for intracranial neural signal. *Advances in Neural Information Processing Systems*, 36:26304–26321, 2023.

Yuchen Zhou, Jiamin Wu, Zichen Ren, Zhouheng Yao, Weiheng Lu, Kunyu Peng, Qihao Zheng, Chunfeng Song, Wanli Ouyang, and Chao Gou. Csbrain: A cross-scale spatiotemporal brain foundation model for eeg decoding. *Advances in Neural Information Processing Systems*, 38: 87150–87195, 2026.

A FROZEN ENCODERS AND PREPROCESSING

We use a lightweight CNN–Transformer autoencoder for self-supervised EEG representation learning, adapted from Ahmadi & Mesin (2025). The encoder contains two temporal convolutional layers, positional encoding, and a two-layer Transformer, producing a 128-dimensional latent representation for each EEG window (1.46M parameters). Following previous studies on EEG representation-learning (Zhou et al., 2026; Jiang et al., 2024; Wang et al., 2025), recordings are band-pass filtered at $0.3 - 75$ Hz, notch-filtered at 60 Hz, mapped to the 19-channel 10–20 montage, resampled at 200 Hz, amplitude-scaled near $\pm 100\,\mu$V, and segmented into non-overlapping 30 s windows. The CNN–Transformer is pretrained on TUEG with signal reconstruction (2,259,998 windows, $\approx$ 18,800 hours). After pretraining, the encoder is frozen and is not updated with TUEP labels or subject identity.

To test dependence on a single representation space, we additionally evaluate LUNA-Base (Döner et al., 2026) (256-d, 7.10M parameters). Table 2 summarizes the two backbones. Frozen-encoder transfer on TUAB and TUEP is reported in Appendix C.1; those numbers are a predictive reference only and do not measure individual specificity.

Table 2: Frozen EEG encoders used for the representation audit. Neither encoder is fine-tuned on TUEP or optimized using subject identity.

Encoder	Parameters	Backbone	Latent dim.
CNN–Transformer	1.46M	CNN + Transformer	128
LUNA-Base	7.10M	LUNA encoder	256

B SIGNATURE DEFINITIONS

Columns of $\widehat{P}$ are source states and rows are destinations. For k=4, $W = \mathrm{vec}(\widehat{P}) \in \mathbb{R}^{16}$ stacks columns:

$$W = \left[\widehat{P}_{11}, \widehat{P}_{21}, \widehat{P}_{31}, \widehat{P}_{41}, \widehat{P}_{12}, \widehat{P}_{22}, \widehat{P}_{32}, \widehat{P}_{42}, \widehat{P}_{13}, \widehat{P}_{23}, \widehat{P}_{33}, \widehat{P}_{43}, \widehat{P}_{14}, \widehat{P}_{24}, \widehat{P}_{34}, \widehat{P}_{44}\right]^{\top}. \tag{11}$$

Residual transitions subtract the occupancy-based expectation $\pi \mathbf{1}^{\top}$ from each column,

$$W^{\perp} = \mathrm{vec}\left(\widehat{P} - \pi \mathbf{1}^{\top}\right), \qquad \widehat{P}^{\perp}_{ji} = \widehat{P}_{ji} - \pi_j. \tag{12}$$

For example, if $\pi_j = 0.40$ but $\widehat{P}_{ji} = 0.65$, the residual is 0.25: the transition is more frequent than occupancy alone would predict.

Dwell is the mean run length in each state. Under an independent sequence the expected run length is $1/(1 - \pi_i)$, so

$$\mathrm{dwell}^{\perp}_i = \mathrm{dwell}_i - \frac{1}{1 - \pi_i}. \tag{13}$$

Self-transition probabilities P_{ii} are the diagonal of $\widehat{P}$. These signatures, together with W, are the temporal probes in Appendix D.

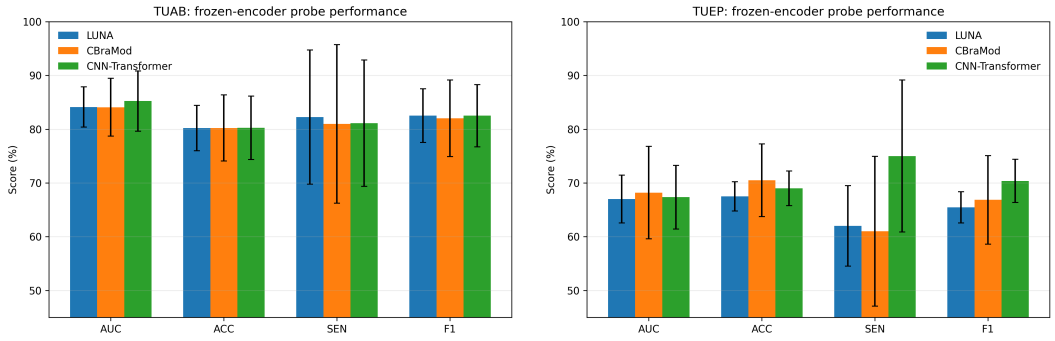
C ADDITIONAL REPRESENTATION ANALYSES

C.1 FROZEN-ENCODER TRANSFER PERFORMANCE

Before examining individual specificity, we evaluated whether the frozen CNN–Transformer representation retained information useful for conventional downstream classification. Under the frozen-encoder protocol, lightweight downstream classifiers were evaluated on the Temple University Hospital EEG Abnormal Corpus (TUAB) and the TUEP without updating the pretrained encoders.

On TUAB, CNN–Transformer achieved an AUC of 85.24 ± 5.61 and an accuracy of 80.25 ± 5.91. On TUEP, it achieved an AUC of 67.35 ± 5.91, sensitivity of 75.00 ± 14.14, and an F1 score of

70.37 ± 4.01. For additional context, we include LUNA-Base and CBraMod as frozen-encoder baselines in this conventional downstream evaluation. CBraMod is included only for the downstream transfer comparison and is not used in the subsequent individual-specificity or temporal-null analyses. Figure 3 provides a visual comparison, with exact values reported in Table 3.



(a) TUAB (b) TUEP

Figure 3: Frozen-encoder downstream performance on (a) TUAB and (b) TUEP. Bars show mean performance and error bars denote standard deviation. CNN–Transformer, LUNA-Base, and CBraMod are compared under the conventional frozen-encoder evaluation. CBraMod is included here as an additional downstream baseline; the individual-specificity audit uses CNN–Transformer and LUNA-Base. These results provide a predictive reference for the frozen representations and do not measure cross-session individual specificity or temporal fingerprintability.

Dataset	Metric	CNN–Transformer	LUNA-Base	CBraMod
TUAB	AUC	**85.24 ± 5.61**	84.13 ± 3.75	84.09 ± 5.39
	ACC	**80.25 ± 5.91**	80.22 ± 4.21	80.23 ± 6.14
	SEN	81.13 ± 11.76	**82.25 ± 12.49**	80.99 ± 14.76
	F1	82.52 ± 5.79	**82.53 ± 5.00**	82.04 ± 7.12
TUEP	AUC	67.35 ± 5.91	67.00 ± 4.45	**68.20 ± 8.60**
	ACC	69.00 ± 3.24	67.50 ± 2.74	**70.50 ± 6.78**
	SEN	**75.00 ± 14.14**	62.00 ± 7.48	61.00 ± 13.93
	F1	**70.37 ± 4.01**	65.46 ± 2.89	66.86 ± 8.23

Table 3: Frozen-encoder downstream classification performance (mean ± standard deviation). The best result for each metric is shown in **bold**, and the second-best is underlined.

Overall, CNN–Transformer performs comparably to the two reference encoders under conventional frozen-encoder evaluation, with the strongest performance depending on the dataset and metric. These results establish that the frozen representation supports downstream prediction, but predictive performance alone does not determine whether it preserves reproducible individual-specific information or whether such information depends on temporal organization. These questions motivate the representation audit in the main analysis.

C.2 Population-level latent structure

As a complementary analysis of the same frozen representation pipeline, we also examined population-level latent structure across clinical groups. Group-associated occupancy changes showed similar patterns across encoders after state alignment. These population-level analyses are distinct from the cross-session individual-specificity experiments that form the primary analysis of this work.

Population-level agreement does not imply that the same latent structure is reproducible within an individual across recording sessions. We therefore evaluate individual specificity separately through within- versus between-subject comparisons in the primary analysis.

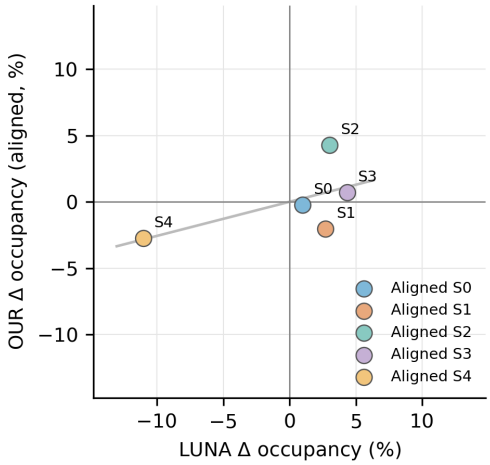


Figure 4: Cross-encoder comparison of group-associated changes in latent-state occupancy after state alignment. Each point represents an aligned latent state. This population-level analysis is complementary to, but distinct from, the cross-session individual-specificity analysis in the main paper.

D Residual and Persistence Diagnostics

Table 4 reports secondary signatures on the same N=112 cross-session pairs. At the primary k=4 setting, none of these signatures shows positive temporal excess ($\Delta_{\text{temp}} > 0$).

Signature	Enc.	Δ_{FP}	Top-1	Perm Δ_{FP}	p_{perm}	Δ_{temp}
Dwell	CNN–Transformer	0.084	0.036	0.093	1.00	−0.009
Dwell	LUNA-Base	0.122	0.040	0.126	0.91	−0.004
dwell$^{\perp}$	CNN–Transformer	0.106	0.058	0.157	1.00	−0.051
dwell$^{\perp}$	LUNA-Base	0.118	0.040	0.194	1.00	−0.076
P_{ii}	CNN–Transformer	0.091	0.045	0.097	0.82	−0.006
P_{ii}	LUNA-Base	0.094	0.031	0.118	1.00	−0.024
$W^{\perp}$	CNN–Transformer	0.089	0.040	0.113	0.95	−0.024
$W^{\perp}$	LUNA-Base	0.098	0.036	0.122	0.91	−0.024

Table 4: Residual and persistence signatures at k=4. Permuted Δ_{FP} denotes the mean over 100 occupancy-preserving temporal shuffles. $\Delta_{\text{temp}} = \Delta_{\text{FP}}^{\text{original}} - \Delta_{\text{FP}}^{\text{permuted}}$.

A dwell-preserving block shuffle reduces W fingerprintability (CNN–Transformer p_{perm}=0.02; LUNA-Base p_{perm}=0), but this null preserves run lengths and therefore does not isolate temporal ordering from occupancy-related persistence.

The occupancy-preserving permutation provides the more direct test of whether the original temporal order contributes individual information beyond occupancy. Under this control, $\Delta_{\text{FP}}(W)$ does not collapse.

Split-half analysis using odd/even recordings (N=91) yields stronger individual specificity, as expected when both halves originate from the same acquisition. For CNN–Transformer, occupancy gives Δ_{FP}=0.228, compared with Δ_{FP}=0.139 for W. However, the occupancy-preserving permutation of W again fails to produce positive temporal excess (p_{perm}=1.0 for both encoders). We therefore treat the split-half results as robustness checks rather than independent evidence for sequential identity.

D.1 Long-recording analysis

To examine whether residual persistence effects remain when transition statistics are estimated from longer recordings, we repeated the analysis using subjects with at least 80 windows in both sessions (N=35).

For CNN–Transformer, residual dwell fingerprintability decreases to Δ_{FP}=0.020 and is not significant (Mann–Whitney p=0.13). This weakening with longer recordings is consistent with the interpretation that apparent residual dwell specificity at shorter sequence lengths can be influenced by finite-sample effects rather than a stable occupancy-independent temporal trait. In contrast, occupancy remains individually informative in this subset (Δ_{FP}=0.150, Top-1 = 0.086; chance = 0.029).

E Cross-Encoder Representation Sensitivity

We additionally asked whether individual structure is preserved across two independently pretrained representation spaces when applied to the *same* EEG recording. This experiment measures cross-representation consistency and should not be interpreted as independent cross-session reproducibility.

For N=198 subjects, Hungarian-aligned mLTD transition signatures at k=4 give a within-subject cosine similarity of 0.518 and a between-subject similarity of 0.447, corresponding to Δ_{FP}=0.071 (p=2.9 $\times$ 10^{-4}). Top-1 retrieval is 0.005, equal to chance. Using held-out state alignment gives Δ_{FP}=0.067 with Top-1 0.033. At k=6, Δ_{FP}=0.046 (p=4.0 $\times$ 10^{-3}) with Top-1 0.018.

These results indicate modest cross-representation preservation of individual structure. Because both representations are derived from the same EEG recording, however, this analysis does not establish cross-session temporal reproducibility. Moreover, the temporal-null experiments in the main analysis show that fingerprintability derived from W should not be interpreted as evidence for unique sequential identity.

F Latent-State Resolution

We examined the within-cluster sum of squares (WCSS) across candidate numbers of latent states for the two frozen encoders used in the individual-specificity analysis. As shown in Figure 5, WCSS decreases monotonically with increasing k for both CNN–Transformer and LUNA-Base, with the largest reductions occurring at small state resolutions.

We use k=4 as the primary low-dimensional state resolution rather than treating it as a uniquely determined optimum. To assess whether the main conclusions depend on this choice, we additionally evaluate neighboring resolutions $k \in \{3, 5, 6\}$ in Appendix G.

For a given k, WCSS is defined as

$$\mathrm{WCSS} = \sum_{i=1}^{k} \sum_{\mathbf{z} \in C_i} \|\mathbf{z} - \boldsymbol{\mu}_i\|_2^2 , \tag{14}$$

where C_i denotes the set of latent representations assigned to state i, and $\boldsymbol{\mu}_i$ denotes its centroid. Lower WCSS indicates greater within-state compactness.

G Sensitivity to the Number of Latent States

Table 5 and Figure 6 repeat the occupancy, transition, and temporal-null analyses for $k \in \{3, 4, 5, 6\}$ using the same N=112 cross-session pairs.

Occupancy fingerprintability varies with state resolution but remains positive across the evaluated values of k. For CNN–Transformer, W does not exceed the occupancy-preserving temporal null at any tested k. For LUNA-Base, p_{perm} remains large at $k \in \{3, 4, 5\}$.

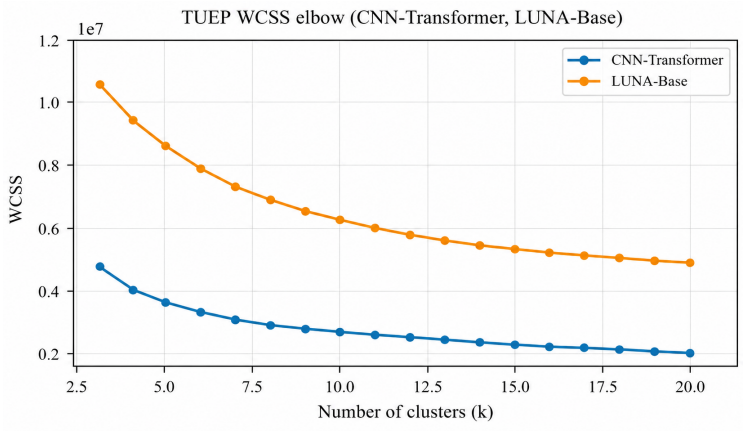


Figure 5: Within-cluster sum of squares (WCSS) across candidate latent-state resolutions on TUEP for CNN–Transformer and LUNA-Base. Both encoders show the largest WCSS reductions at small values of k. We use k=4 as the primary low-dimensional state resolution and evaluate $k \in \{3, 5, 6\}$ as sensitivity settings.

k	Enc.	Occ. Δ_{FP} (Top-1)	W Δ_{FP} (Top-1)	W permute Δ_{FP}	p_{perm}
3	CNN–Transformer	0.058 (0.022)	0.067 (0.031)	0.070	0.70
3	LUNA-Base	0.052 (0.009)	0.054 (0.036)	0.060	0.82
4	CNN–Transformer	0.103 (0.018)	0.106 (0.058)	0.108	0.63
4	LUNA-Base	0.096 (0.009)	0.095 (0.045)	0.105	0.92
5	CNN–Transformer	0.092 (0.049)	0.060 (0.027)	0.080	1.00
5	LUNA-Base	0.078 (0.022)	0.085 (0.045)	0.079	0.23
6	CNN–Transformer	0.098 (0.040)	0.072 (0.040)	0.081	0.91
6	LUNA-Base	0.087 (0.040)	0.084 (0.058)	0.070	0.03

Table 5: Cross-session sensitivity to latent-state resolution (N=112; chance Top-1 = 0.009). Permute denotes 100 occupancy-preserving shuffles of lag-1 W.

LUNA-Base at k=6 is the only secondary setting with p_{perm}=0.03, corresponding to $\Delta_{\mathrm{temp}} = 0.014$. Because this small positive excess occurs in one secondary state resolution and is not replicated by CNN–Transformer, we treat it as sensitivity evidence rather than evidence for a robust temporal fingerprint. The pre-specified k=4 analysis therefore remains the primary result.

H Secondary group comparison of signatures

The main text does not use diagnosis for clustering, signature estimation, or the temporal-null test. As a post-hoc check we compare the same k=4 signatures between epilepsy and non-epilepsy on all labeled subjects (n=99/99), pooling recordings within subject. TUEP non-epilepsy is a clinical contrast, not a healthy-control cohort. Table 6 reports mean pairwise cosine; Table 7 reports Mann–Whitney tests on scalars. Epilepsy subjects are slightly more similar to each other than non-epilepsy subjects, but epi-versus-non cosine remains high. The strongest scalar gaps are mean self-transition P_{ii} and occupancy entropy. Residual $\|W^{\perp}\|_F$ does not differ. This is a population stickiness / occupancy-mix effect, not sequential identity, and it does not change $\Delta_{\mathrm{temp}}(W) \leq 0$.

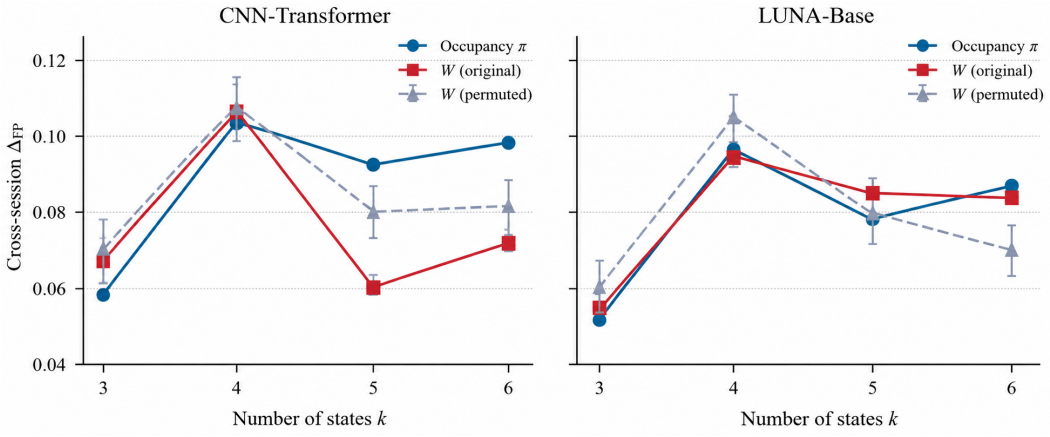


Figure 6: Sensitivity of cross-session Δ_{FP} to the number of latent states (N=112). Original W generally tracks the occupancy-preserving permutation across state resolutions. LUNA-Base at k=6 is the only sensitivity setting with a small positive original–permuted excess; because this effect is not replicated by CNN–Transformer or at the primary k=4 setting, we treat it as a sensitivity result rather than a primary finding. Error bars denote SD over 100 occupancy-preserving permutations.

Sig.	Enc.	Among epi	Among non	Epi vs. non
π	CNN–Transformer	0.605	0.492	0.541
π	LUNA-Base	0.661	0.570	0.603
W	CNN–Transformer	0.625	0.455	0.532
W	LUNA-Base	0.664	0.519	0.568
$W^{\perp}$	CNN–Transformer	0.455	0.287	0.354
$W^{\perp}$	LUNA-Base	0.503	0.357	0.387
Dwell	CNN–Transformer	0.675	0.555	0.613
Dwell	LUNA-Base	0.683	0.604	0.614
P_{ii}	CNN–Transformer	0.726	0.560	0.638
P_{ii}	LUNA-Base	0.763	0.613	0.662

Table 6: Mean pairwise cosine of subject-level signatures (k=4, n=99/99).

Scalar	Enc.	Epi	Non	MW p
Mean P_{ii}	CNN–Transformer	0.531	0.408	7.1×10^{-7}
Mean P_{ii}	LUNA-Base	0.546	0.401	1.1×10^{-8}
Occupancy entropy	CNN–Transformer	0.912	0.753	7.3×10^{-4}
Occupancy entropy	LUNA-Base	1.010	0.732	9.3×10^{-8}
$\|W\|_F$	CNN–Transformer	1.463	1.379	4.5×10^{-3}
$\|W\|_F$	LUNA-Base	1.485	1.356	1.8×10^{-5}
Mean dwell	CNN–Transformer	3.94	3.20	0.011
Mean dwell	LUNA-Base	3.64	3.38	0.014
$\|W^{\perp}\|_F$	CNN–Transformer	1.179	1.128	0.24
$\|W^{\perp}\|_F$	LUNA-Base	1.098	1.149	0.24

Table 7: Epilepsy vs. non-epilepsy scalars (k=4, n=99/99). Mann–Whitney two-sided p.