Approaching models of (sub)culture: SWAG

Pierre-Louis Suckrow, Suzan Hanow

suckrowpierre@gmail.com, suzanelaa@gmail.com

Abstract

This study examines contemporary artificial intelligence models not as simulations of human cognition, but as representations of cultural patterns, specifically using the aesthetic subculture of SWAG as a case study Utilizing a transdisciplinary methodology that blends media theory, cognitive science, and information aesthetics, we introduce a new framework for analyzing subcultural phenomena within the realm of computational intelligence. In conjunction with this framework, we develop a SWAG Measuring Algorithm using state-of-the-art text-image embedding models. Further building on this foundation, we employ dimensionality reduction techniques combined with methods from archaeology to create a digital probing tool. This tool enables us to examine how embedding models interpret subculture and to identify potential biases inherent in those interpretations. To assess the effectiveness of various embedding models in representing subcultural elements, we designed and conducted a study that established a human-annotated baseline. This study also uncovered biases present both in the models and among the human participants. The findings from our research and study are presented in an interactive multimedia installation. The resulting exhibition pieces aim to encourage viewers to shift their perspective—from viewing artificial models as mere expressions of intelligence to recognizing them as culturally situated systems of representation.

1. Introduction

Modeling intelligence has been the hottest topic these past years. The rate at which so-called “artificial models” have been released has been steadily increasing. Largely driven by the commercialization and marketing of these models, the primary emphasis has been on achieving high performance on benchmarks and maximizing profitability, often suggesting a proximity to artificial general intelligence (AGI). As these models increasingly permeate everyday life, the collective understanding of these models has been that they are inherently intelligent. However, these models are nothing more then reductionist data processing, stochastic predictions, and smart pattern recognition.

The conceptual foundation of such models can be traced back to Frank Rosenblatt’s development of the perceptron (Rosenblatt, 1958), which laid the groundwork for contemporary deep learning models capable of tasks such as multi-modal generation and classification. These capabilities arise from training on extensive amounts of task-specific data. For instance, a model trained on a vast corpus of textual data can generate text that appears coherent and contextually appropriate. However, it is crucial to recognize the origins of this data. Image, text, audio and any other media corpora only exists because of culture and in manifestation in artefacts and collective memory/history.

By reframing contemporary deep learning models not as models of intelligence but rather as *Models of Culture*, a novel perspective emerges for examining, analyzing, and understanding cultural phenomena (Underwood, 2021). Our work builds upon this conceptual approach to introduce new frameworks and methodologies for the analysis, investigation, and communication of (sub)cultural phenomena. Specifically, it applies the *Models of Culture* perspective to the subculture of *SWAG*, allowing to investigate how selected contemporary “artificial models” interpret and represent *SWAG*-culture. These new frameworks and methodologies were developed while addressing key questions regarding the understanding of *SWAG*-culture in artificial models while also presenting findings through multiple artistic exhibition pieces.

Is SWAG measurable? - As of the time of writing SWAG is on the verge of coming back (HeyJodye, 2025; Kroening, 2025). Our culture at large seemingly lives in a circle of reemergence of past trends. Reliving memories and culture through the romanticised lense of nostalgia, we express the urge for seemingly better times in our enactment of the current Zeitgeist trough reinterpretation and remixing of artefacts specific to this timeframe. SWAG, Bling Culture and Indie Sleaze were eras at the end of the 2000 /beginning of the 2010 defined by specific style cues and elements. Crucial elements of this culture were being captured and expressed across mediums such as music, videos, photos, memes and fashion all being archived and presented on early social media channels and the internet at large. Being inspired especially by fashion we turned to a task that would reflect the reductionistic design of machine learning. The evaluation of a person style. Applied to the subculture of swag, we obtain a measurement of the SWAG of a person. On might assume that such a task might be very hard to define and operationalize due to the very vague nature of the parameters like clothing, accessories, and attitude and character leading to the same inconclusive results as the field of Information Aesthetics. But unlike Max Bense and Abraham A. Moles the framing through models of cultures allows us to try to map these very elusive concepts in a very elusive tool itself. *Embedding spaces.*

Consequently, our first research question emerges: **RQ 1: Can embeddings and latent spaces be used to measure the styled basesd asthetics value (swag) of a person?**

How accurate is this measurement? To our surprise, the application of embedding spaces to the concept of *SWAG* demonstrated high effectiveness without requiring additional detailed descriptions, training, or fine-tuning. To assess this, we evaluated three different text/image embedding models based on the *CLIP* architecture. To validate our initial subjective impression, we designed a study to systematically evaluate multible CLIP models by establishing a baseline for comparison.

This study serves not only to establish a baseline but also aims to address the following second research question: **RQ 2: What cultural biases and patterns can be observed in the latent understanding of Swag?**

How does this understanding come from? - Interpreting the internal mechanisms of deep machine learning models is inherently difficult. Especially the complex structure of embedding spaces limits our ability to understand their reasoning, which is critical for explaining decisions, debugging, comparing alternatives, and identifying hidden biases (Simhi & Markovitch, 2023). To enhance interpretability within our context, methods such as conceptualization, dimensionality reduction, and sampling were employed, culminating in a **Digital Probing Device**.

How do we communicate this shift in perspective of "artificial models"? Beyond merely illustrating the reductive processes underlying machine learning, our principal objective includes facilitating a collective transition from perceiving these systems merely as models of intelligence toward recognizing them as models of culture. The presented exhibition artifacts exemplify this by demonstrating how projections of high-dimensional latent spaces into three dimensions offer a partial yet insightful perspective on their internal logic, while also bringing awareness to the underlying cultural dynamics being analyzed.

2. Background

Memory Artefacts - The Extended Mind Thesis, introduced by Clark and Chalmers (1998), argues that cognition extends beyond the confines of the brain and body, encompassing external environments and the objects with which individuals interact. According to this perspective, external tools and objects play a crucial role in cognitive processes related to memory. Throughout history, humans have consistently relied on external means for storing and transmitting information, evolving from cave paintings and language to contemporary computational systems. Merlin Donald describes this external storage system, referred to as *exograms*, as a fundamental feature distinguishing human evolution and consciousness (Donald, 1993; Heersmink, 2024). These exograms function as *memory artefacts*: objects or structures explicitly designed to facilitate memory-related activities. They assist users in recalling and conveying experiences, events, facts, or other forms of information, thus supporting individual memory tasks and acting as conduits for cultural transmission (Heersmink, 2024). Consequently, these artefacts play a vital role in maintaining and communicating cultural heritage and individual identities across generations.

Media and Culture — Media serve as essential instruments in shaping cultural identity and collective memory, functioning both as vessels of transmission and as frameworks for interpretation. According to Assmann and Hölscher (1988), collective memory is not biologically inherited but culturally and socially constructed, enabling societies to sustain a shared historical consciousness across generations. He distinguishes between communicative memory, which is limited to interpersonal interaction over a few decades, and cultural memory, which gains durability through media, rituals, and institutional preservation. In this context, media act not only as repositories but also as active constructors of cultural memory, selectively recording and reconfiguring historical narratives. The role of digital platforms has expanded this function significantly: through images, videos, and other media forms, social media embeds memory into everyday digital practices and integrates past and present. In recent years, this process has become increasingly shaped by algorithmic systems, which influence what is remembered, circulated, and emphasized. Especially in subcultural contexts, digital media have assumed a central role in expressing identity, enabling participation and stylistic differentiation across global networks. Aesthetic and symbolic elements—once mediated through physical formats such as magazines or underground broadcasts—are now recontextualized within visual platforms like Tumblr, YouTube, and Instagram. These platforms not only archive cultural practices but also modulate their visibility and significance, thereby contributing to the continuous negotiation of cultural meaning in a networked media environment.

Information Astehtics - The question of how aesthetics can be evaluated has long intrigued scholars, motivating researchers such as Max Bense and Abraham A. Moles to establish the field of Information Aesthetics during the 1960s. Drawing from Claude E. Shannon's foundational work in Information Theory (Shannon & Weaver, 1998), their objective was to devise a mathematical framework for objectively quantifying the aesthetic quality of an object or concept. While these initial approaches were intellectually stimulating and provocative, they received limited attention and eventually receded into obscurity. The inherently reductionist and schematic nature of these methodologies aligns closely with the operational principles of contemporary machine learning models, which are similarly characterized by reductionist and schematic representations. (Nake, 2012)

Generative AI - "One of the most distinctive and powerful aspects of human cognition is the ability to imagine: to synthesize mental objects which are not bound by what is immediately present in reality (...). The subfield of Machine Learning which aims to endow machines with this same essential capacity to imagine and synthesize new entities is referred to as generative modeling" (Lamb, 2021, p. 1). Synthesizing data via machine learning models is rendered possible under the assumption that examined data originates from a specific underlying distribution. Gm et al. (2020), Lamb (2021), and Ruthotto and Haber (2021) describe that the objective of such models is to approximate this high-dimensional probability distribution during training, thus enabling the extraction of novel data points through distribution sampling at inference time.

Embeddings - One of the most commonly used methods for computational translation and interpretation of media involves representing it as vectors. This process, referred to as *embedding*, creates an n-dimensional space termed the *latent space*. This mapping is learned through training on data, resulting in a vector space where semantic and other meanings are encoded (Levy & Goldberg, 2014). Consequently, related vectors share similar directions and are positioned closer to each other compared to unrelated words. This concept of distance can also be employed to translate the relationship between one pair of words to another, thereby identifying the appropriate word that aligns with the given semantic correlation. For instance, by taking the distance between the words "Germany" and "Japan" in the high-dimensional space and adding the distance to the word "Sushi," one would find "Bratwurst" near the new resulting position (3Blue1Brown, 2024). From this notion of semantics models learn to recognize statistical patterns while also taking into account different relations.

Subjectivity in computational models - Being concieved from mathematical theorems and principles and governed by formal logic computation has an inherent objectivity to it. This holds true for its most atomic operations but the layers of abstraction introduced by humans to make its understanding and operation more intuitive also intoduced subjectivity. This is especially relevant in models of language, due to the subjectivity of natural language itself Baeyaert, 2024. Large Language Models (LLM)s and current modelings of intellgience draw from data formed by human culture, absorbing and reflecting, these intrinsic human biases and perspectives.

Models of Culture - Underwood (2021) proposes shifting from viewing artificial intelligence models primarily as abstraction of individual cognition to understanding them as representations of cultural practices. While current artificial intelligence models cannot yet capture individual intentions or mental states, defining meaning strictly as intentional communication between individuals neglects broader perspectives common in historical disciplines. Historians typically interpret meaning as collectively constructed rather than tied exclusively to individual intent. They locate meaning within broader cultural patterns such as literary genres, political movements, or traditional motifs. Therefore, artificial intelligence models hold significant value not because they replicate individual thought processes, but because they embody culturally specific patterns—like styles, discourses, or narrative structures—that scholars can analyze, compare, and creatively reinterpret. Unlike general intelligence models, cultural models are historically specific and inevitably contain biases and omissions reflective of their contexts. Thus, reframing artificial intelligence as models of culture allows historical disciplines to better utilize these technologies to identify differences and changes across periods, sources, or communities This perspective emphasizes both the potential and limitations of artificial intelligence in exploring, interpreting, and comprehending cultural phenomena through computational means.

Conceptualization - To enhance the interpretability of embedding spaces, Simhi and Markovitch (2023) introduces a method that projects latent embeddings into a human-readable conceptual space. This approach relies on a predefined set of textual descriptions representing various concepts within the space. The embedding of each concept in the investigated model is computed in advance. When processing a new input text, the method evaluates its similarity or distance to each concept's embedding, thereby providing a more interpretable representation of the original embedding and the latent space. This technique facilitates the analysis of how the model encodes semantic information, enabling the identification of potential biases and serving as a debugging tool. Furthermore, it supports the comparison of different embedding spaces by highlighting variations in their semantic encoding structures.

Dimensionality Reduction - Visualization and classification are essential components of exploratory data analysis. Dimensionality reduction helps in understanding complex datasets by transforming high-dimensional data into lower-dimensional spaces. This process focuses on identifying and retaining the most important features while reducing or combining less relevant information Engel et al. (2013). Dimensionality reduction serves different purposes: for visualization, while for classification tasks, the goal is to produce mappings that separate distinct categories within the data (Vlachos et al., 2002). Ideally, the distances and spatial relationships among points in reduced-dimensional representations should accurately correspond to their original relationships. Due to the inherent complexity of high-dimensional data, exact preservation of these relationships is often impossible, leading to projections that inherently carry ambiguity Engel et al. (2013). Nonetheless, dimensionality reduction effectively captures users' perceptions of similarities in high-dimensional data, significantly improving the capability for meaningful visualization (Vlachos et al., 2002).

2.1 Reflection on selection of references

Underwood's *Models of Culture* (2021) was central in shaping our theoretical framework. His proposition to interpret AI systems not as models of cognition, but as models of culture, provided a productive shift in perspective—one that aligned with our interest in investigating SWAG not as a fixed aesthetic category, but as a dynamic, culturally situated phenomenon. Similarly, historical perspectives on Information Aesthetics, particularly those of Max Bense and Abraham A. Moles, offered a conceptual bridge between the quantification of aesthetics and the computational logic underlying latent spaces. While we do not adopt their frameworks uncritically, their work provided a historical lens through which we could reflect on our own methodological choices.

The inclusion of media theory and memory studies—particularly the writings of Jan Assmann, Heersmink, and Donald—served to contextualize our understanding of fashion artefacts as carriers of cultural memory. These sources emphasized the role of externalized forms of knowledge (exograms, digital archives, artefacts) in the construction and transmission of collective identity. This was particularly relevant in our attempt to understand how SWAG is encoded in and reinterpreted by machine learning models.

3. Methodology

Originally focused on critically exploring computational intelligence, psychology, and visual communication, our research unexpectedly evolved to encompass fashion and its sociocultural dimensions. This shift became a tool for critically reflecting on design patterns and the implications of the expanding use of computational intelligence.

This transition occurred after an extensive research and literature review phase, during which we examined topics related to memory, culture, psychology, and symbolism in search of a relevant context to connect these themes to our current engagement with computational intelligence. This transdisciplinary reframing, viewed through the lenses of archival practices, fashion, and information aesthetics, allowed us to articulate our initial ideas more comprehensibly while introducing a playful element to our work.

3.1 Measuring SWAG

Searching for a channel to critically reflect upon and reorient the prevailing application and understanding of computational intelligence models, we turned our attention to the fashion domain.

Wanting to communicate the reductionist design principles inherent in computational intelligence models, we choose to model an analogous approach through them: objectifying/quantifying personal style. Fashion is not just about clothing; it is an intersection of multiple contextual elements, such as personal lifestyle, historical references, and socio-cultural identity. A person's outfit choice can reflect affiliations with specific fashion subcultures, adherence to current trends, or a rejection of conventional styling norms. However, capturing all these dimensions computationally in their full complexity would be impractical. Instead, computational intelligence models might abstract these intricate layers of meaning, filtering out redundant or less influential factors and prioritizing the most visually and contextually significant attributes.

Introducing the concept of *models of culture* into this framework enables the integration of a shift in our understanding of computational intelligence models. To achieve this, we include a subcultural layer to the quantification of personal style. Selecting a well-documented "narrow" subculture within internet culture makes it more likely to be represented in the training data and embedding spaces of contemporary computational intelligence models. Aiming to adopt a playful element that might capture attention, we chose SWAG culture as our subcultural layer, reframing the quantifying personal style as the quantification of an individual's SWAG.

The quantification of personal style can be understood as a quantification of aesthetics, establishing a connection to Information Aesthetics. Utilizing embeddings to estimate an aesthetic value represents an approach not previously accessible to researchers in this field. While embeddings provide a powerful tool for encoding complex and difficult-to-describe parameters, they are not inherently objective, an aspect that aligns with the foundational principles of Information Aesthetics. The use of embeddings in our quantification is not truly "objective" but rather a subjective measure influenced by the choice of models, their training data, training objectives, learning algorithms, and their interpretation within our methodology. Despite this, this approach still facilitates a quantifiable assessment of SWAG by systematically mapping subjective and culturally influenced perceptions of fashion onto a structured computational framework.

To digitally quantify the SWAG of a person's outfit, evaluation or prediction can be performed on an image depicting the individual or the outfit. Embeddings enable the transformation of this image into a representation suitable for computation, comparison, classification, and analysis, incorporating cultural context and semantic information. The embedding model must contain associations with the term "SWAG" or related cultural references to perform this analysis within the embedding space. Consequently, the model requires an understanding of cultural nuances and visual cues linked to the SWAG, restricting the model selection to ones capable of embedding images and text within a unified latent space.

3.1.1 Measurment via CLIP

One of the most widely adopted architectures for text-image embedding models is CLIP (Radford et al., 2021), a neural network designed to learn visual concepts by associating images with corresponding textual descriptions sourced from the internet. This adaptable model allows fine-tuning or complete retraining using diverse datasets. For this study, we selected five distinct models for comparative analysis. Initially, we employed two models published by OpenAI: *openai/clip-vit-base-patch16*[1] and *openai/clip-vit-large-patch14-336*[2]. Considering the assumption that a model fine-tuned specifically for fashion-related tasks would yield improved performance for our application, we also included the *patrickjohncyh/fashion-clip*[3] model. Finally, we selected the currently most popular CLIP-based model on Huggingface, *laion/CLIP-ViT-bigG-14-laion2B-39B-b160k*[4] and its smaller version *laion/CLIP-ViT-B-32-laion2B-s34B-b79K*[5].

We employ a standard classification approach to measure an individual's SWAG using one of the selected models. This

[1] `https://huggingface.co/openai/clip-vit-base-patch16`

[2] `https://huggingface.co/openai/clip-vit-large-patch14-336`

[3] `https://huggingface.co/patrickjohncyh/fashion-clip`

[4] `https://huggingface.co/laion/CLIP-ViT-bigG-14-laion2B-39B-b160k`

[5] `https://huggingface.co/laion/CLIP-ViT-B-32-laion2B-s34B-b79K`

method provides the model with a set of textual descriptions or concepts, including the keyword "SWAG". The model calculates similarity scores between the image embedding and each provided textual description through the dot product when presented with an image. These computed similarity scores are subsequently normalized into a probability distribution, enabling quantification of SWAG in percentages.

This methodology developed for measuring SWAG is influenced by two main factors. Firstly, the effectiveness of this measurement depends on the embedding model's ability to comprehend the concept of SWAG, which relies on its training data and learned associations. Secondly, accuracy is significantly impacted by the selection of categories (concepts) chosen for the classification task. Selecting overly similar concepts may lead to ambiguous predictions, while entirely unrelated concepts can increase false-positive classifications of SWAG. Additionally, the total number of concepts chosen directly affects the probability distribution outcomes. Achieving the highest accuracy requires careful optimization of these factors.

3.2 Digital Probing Device

Surprised by how effectively some embedding spaces captured the subjective concept of SWAG without requiring detailed descriptions or additional training, we sought to investigate how this understanding developed. Drawing inspiration from archaeological methodologies, we developed an approach to investigate the shared text-image embedding space. By employing dimensionality reduction techniques, we enable interactive exploration and visualization of emergent clusters and structural relationships within a three-dimensional space.

Archaeology is the study of human cultures by recovering, analyzing, and interpreting artifacts and material remains (Smith, 2014). Substituting the investigated subject with our current modeling of intelligence/culture and shifting material artifacts to a digital realm, we obtain an isomorphic approach for exploring how computational intelligence models and high-dimensional spaces might operate.

The method of probing in archaeology refers to the technique of employing a probe to penetrate the ground or surface to investigate concealed layers or structures in the hope of making contact with target artifacts or archaeological features (Darvill, 2010). Analogously, we employ this probing methodology to analyze dimensionality-reduced embedding spaces, thereby facilitating the exploration of underlying structures and conceptual relationships. Reducing the embedding space to three dimensions translates complex, high-dimensional spatial relationships into a comprehensible visual form. We specifically apply *Principal Component Analysis* (*PCA*) to both the images being evaluated and the textual representation of the concept SWAG. The resulting three-dimensional embedding space is normalized to fit within a unit cube bounded by $[-1, 1]^3$ (see figure 1). Within this space, the word "SWAG" is positioned at the center, at the coordinates $(0, 0, 0)$. This central location provides a reference point for comparing the embeddings of the images, aiding in the understanding of their semantic relationships. Consequently, images corresponding closely to visual concepts of SWAG are situated near the origin, whereas images less aligned with these concepts are positioned closer to the boundaries of the cube. By using a virtual probing device movable within the unit cube, data points proximate to its current position can be shown.

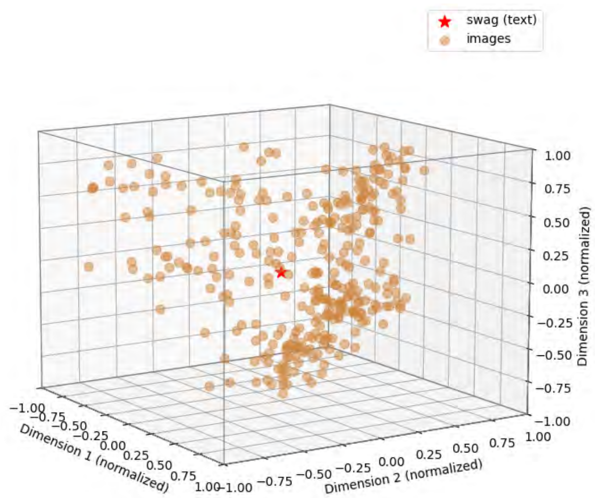

Figure 1

CLIP-based 3D PCA mapping of 316 images relative to the text prompt "swag". The red star at the origin represents the normalized text embedding, while the points correspond to the image embeddings reduced in dimensionality.

Due to the normalization process, the careful selection of data points for dimensionality reduction is essential. Selecting excessive "SWAGGY" images may compromise the concentration around the center. Conversely, selecting an insufficient number of these images might fail to capture structural information and patterns associated with SWAG. Therefore, ensuring a balanced representation of various styles and individuals within the dataset is crucial to minimize bias in the probing methodology.

3.3 Study Design

To evaluate the performance of selected embedding models in predicting SWAG, we established a human-annotated baseline through a controlled user study. The goal of this study was to gather subjective ratings of SWAG across a diverse set of images. We created a dedicated website to facilitate data collection, in which participants were asked to rate the overall impression of SWAG conveyed by each image, considering not only clothing but also body language and physical presence. The ratings were recorded on a continuous scale from 0 to 100, where 0 indicated a complete absence of SWAG and 100 represented the highest level of perceived SWAG.

The used image dataset was curated to encompass a combination of stylistic and demographic attributes. Each image was classified into one of three style categories (swaggy, basic, corporate), one of two gender presentations (male, female), one of three ethnic groups (People of Color, Caucasian, Asian), one of two age groups (young adult, adult), one of two framing types (upper body, full body), and one of two pose conditions (posed, non-posed). These six dimensions of attributes yielded a total of ($3 \times 2 \times 3 \times 2 \times 2 \times 2 = 144$) unique category combinations. Two distinct images were selected for each combination, resulting in 288 images.

The study involved $n = 40$ participants with an average age of 24. 20 of the participants were male, 19 female and 1 person identfied as other. Each participant was tasked with rating a fixed number of images, sampled to ensure balanced coverage and redundancy. The target was to obtain 10 independent ratings for each image, yielding $288 \times 10 = 2880$ total required ratings. Consequently, each participant rated 72 images.

In addition, 28 supplementary images were included in the study. These images, while still related to the concept of SWAG, incorporated humorous elements. Their purpose was to introduce variation and mitigate potential fatigue associated with the continuous evaluation of standard images.

4. Process

4.1 Project Reflection and Process Narrative

Our project began by exploring the relationship between generative artificial intelligence and memory, with particular attention paid to the role of memory artifacts in training data. In the early stages, the investigation focused on the potential implications of generative AI for individual processes of memory and recollection. As the research progressed, attention shifted toward the concept of collective memory, leading to theoretical engagement with the work of Jung et al. (1968) and his notion of a shared unconscious. At this juncture, we remained uncertain regarding the ultimate outcomes of our research.

Feeling stuck, we conducted a workshop together with student peers of the master program “Design and Computation”, with the goal of acquiring new perspectives and challenging initial assumptions. The central question regarding the relevance of the main themes formed the basis for a broader discussion that led to topics such as archival practices, fashion, and cultural practices. While no definitive resolution came from this exchange, it provided an important space for critical dialogue where we could reflect on and reorient our emerging ideas.

An initial, somewhat humorous proposal to employ computational models to evaluate “SWAG” evolved into a more nuanced investigation of whether embedding spaces could be utilized to analyze the aesthetic and stylistic properties of subcultural expression. Surprisingly, this approach yielded coherent and interpretable results. The models appeared to capture implicit cultural signals and stylistic cues with a degree of specificity that exceeded initial expectations. This observation prompted a further reorientation of the project, shifting the focus to the mechanisms through which such semantic and aesthetic understandings are encoded and operationalized within model architectures.

By creating a narrative that incorporates our core ideas with key insights, we reduced our research into a conceptual framework, which was then expressed in both theoretical and practical formats. We developed three exhibition pieces to showcase the project’s main insights. Additionally, we designed an empirical study to evaluate the performance of various embedding models on the proposed task and to assess any potential biases.

4.2 Design Artifacts

The Swagometer (see figure 4.2) is our first developed prototype, designed as an application for the iPhone 4s, which serves as the user interface. We purposefully selected this outdated device to evoke the period around 2013–2014, often considered the height of the “Swag” era, tapping into a sense of nostalgia. The application enables users to take a photo of themselves or others and generates a percentage score reflecting their "SWAG," based on our proposed method. The central idea of this exhibit is to engage the audience while also demonstrating the reductionist tendencies of computational intelligence by reducing visitors down to a singular SWAG value.

Expanding upon our initial exhibition element, we developed a second exhibition element to engage with the digital probing feature. This setup consists of two monitors, with the first displaying the projection cube featuring the dimensionality-reduced images (see figure 4.2). Visitors can navigate and manipulate the cube using a provided mouse to control the camera and a MIDI joystick (see figure 4.2) to rotate the cube, which can only be turned up or down or left or right.

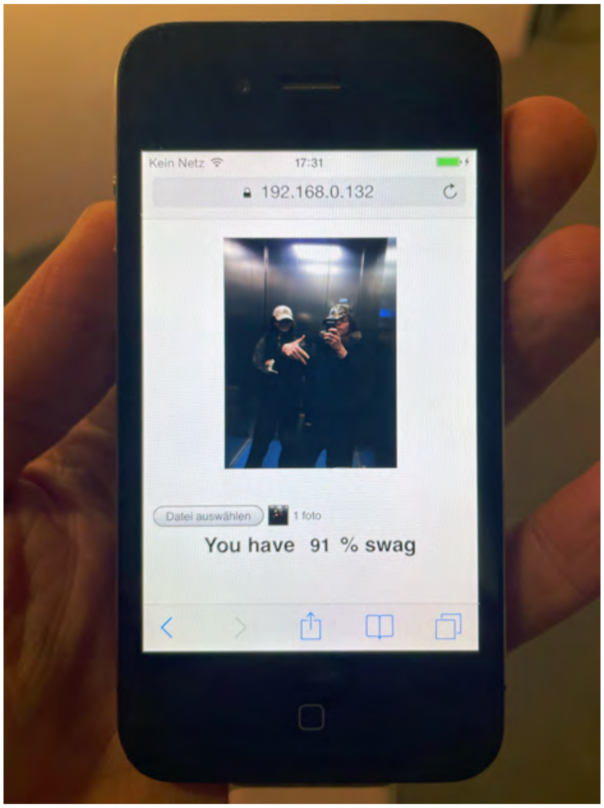

Figure 2
Swagometer

In this 3D environment, a cylinder serves as the probing device. Visitors can move this probing device closer to or further away from the cube by using the buttons on the joystick. When the probing device is inserted in the cube and in proximity to a data point, that data point is displayed on the second monitor. If the joystick remains inactive for a certain duration or if the probing device is outside the cube, the second monitor will present text that describes the methodology and background of the project (see figure 4.2), as well as instructions for interacting with the device.

In our last exhibition element, we aimed to communicate how latent spaces capture culture and how we use dimensionality reduction as a tool to make these abstract spaces more understandable. To achieve this, we constructed a cube made of a metal frame covered with semi-transparent, latex-textured polyurethane (PU) layers (see figure 6). These layers allowed color and light to shine through while obscuring any objects placed inside the cube. The bottom of the cube was left open and rested on a pedestal, which housed multiple projectors directing images onto the surfaces of the cube.

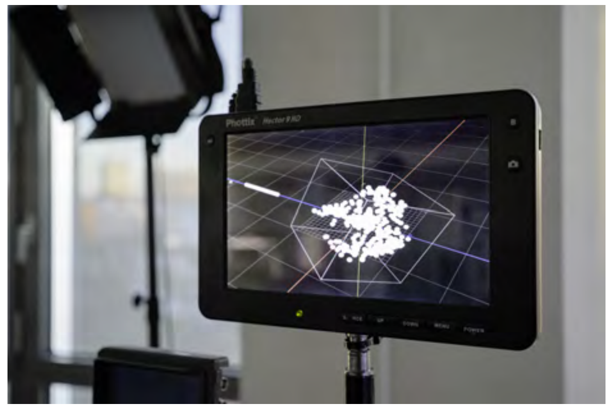

Figure 3
Monitor 1

Figure 4
Controller

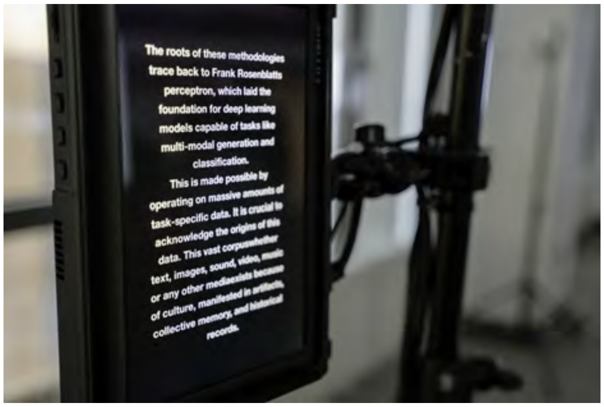

Figure 5
Monitor 2

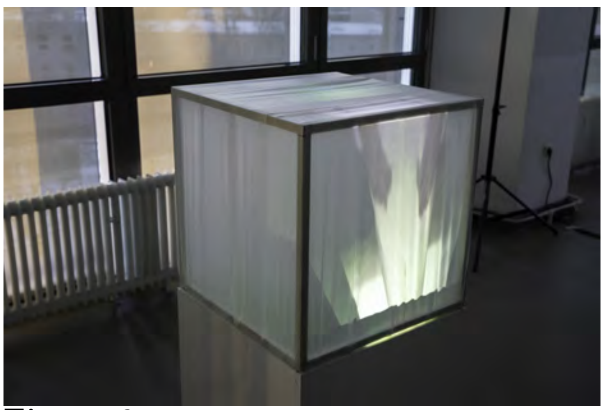

Figure 6
Projection Cube

This cube represents the dimensionality-reduced latent space, with the projectors symbolizing the latent space itself. Various videos projected onto the surfaces illustrate different aspects of the analyzed subculture. This setup provides a tangible outer shell that makes otherwise incomprehensible high-dimensional representations accessible through projection. As participants manipulate or remove these layers of abstraction, the projected light escapes, revealing the underlying complexity of high-dimensional space. Without these layers, the high-dimensional representation of culture becomes unintelligible, highlighting the crucial role of abstraction in making complex cultural models understandable.

5. Findings and Results

5.1 Swag Measurment and Digital Probing Device

By constructing a conceptual framework that spans multiple disciplines —including psychology, cognitive science, cultural science, computer science, information aesthetics, media science, fashion, and archiving —we demonstrated how current models of intelligence can be utilized as tools for interacting with and analyzing (sub)cultural practices and phenomena.

We achieved this by using these models to quantify the aesthetic values of a specific subculture, namely SWAG. Additionally, drawing inspiration from investigative methods in archaeology, we developed a refined and interactive visualization of the inherent modeling within these intelligence and culture models while also raising awareness of the model rendering of this culture.

5.2 Study Results

Using the collected data, we calculated baseline values for our analysis, including the mean and median ratings along with their confidence intervals.

For the mean rating's confidence interval, we determined the standard deviation and the number of valid ratings. The standard error was calculated by dividing the standard deviation by the square root of the rating count, and the 95% confidence interval was derived using a z-factor of 1.96.

We also calculated the median rating's confidence interval using a bootstrap method, resampling the data to approximate its distribution. The interval was defined by the 2.5th and 97.5th percentiles of this distribution.

Our data's high standard deviation and variance indicate significant variability in the ratings. A subset of "high-confidence" images was identified, defined as those with a 95% confidence interval width under 30 and a standard error below 5. Out of 288 evaluated images, only 60 met these criteria, suggesting that most ratings exhibited considerable variability and lower precision.

Apart from the inherently subjective nature of the evaluation task, another potential contributing factor to the observed variability is that some participants appeared to rate the outfits based on personal preference rather than specifically evaluating the level of SWAG depicted in each image.

After establishing baseline measurements, we compared the performance of different models. Each model assessed SWAG as described in section 3.1.1, utilizing three distinct concept lists, resulting in a total of 5×3 model-concept combinations. We calculated the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to measure deviations from the baseline. Statistical tests, including a paired t-test for mean ratings and a Wilcoxon signed-rank test for median ratings, assessed the significance of these differences, yielding p-values for discrepancies between model predictions and human baselines. The evaluation included the full image dataset (see Appendix, Table A2) and a high-confidence subset of 60 images with tight confidence intervals (see Appendix, Table A3), allowing for both overall performance assessment and focused analysis of precise ratings.

The interpretation of the obtained results can be summarized as follows: models with p-values below 0.05 show significant errors compared to the baseline and are unsuitable. Lower MAE and RMSE values, derived from median and mean ratings, signify better model performance against baselines. The model *openai/clip-vit-base-patch16* with the concepts swag, no swag consistently achieved the lowest error values across both comprehensive and high-confidence evaluations. It was followed closely by the same model and *openai/clip-vit-large-patch14-336* using swag, sportive style, corporate style, casual style, elegant style, unfashionable in high-confidence evaluations. The model *openai/clip-vit-large-patch14-336* with wag basic corporate and *patrickjohncyh/fashion-clip* with the same concept set also performed well in all evaluations.

Calculating the average SWAG ratings for both the models and the participant study across different attributes revealed certain tendencies and biases regarding the understanding of SWAG. Participants rated male images about 10% higher in SWAG than female images, while Laion CLIP models rated females up to 30% higher. Participants assigned nearly double the SWAG to people of color compared to Caucasian and Asian individuals, a trend seen to a lesser extent in models. Ratings for Caucasian and Asian individuals were similar across groups. Participants favored younger individuals in SWAG ratings, whereas models showed the opposite trend. Overall, the impact of full-body versus upper-body images was minimal, and while posed images were favored in model evaluations, this trend was not reflected in participant ratings.

6. Discussion

Our results show that embedding spaces using CLIP-based architectures, can effectively measure subjective aesthetic values related to the SWAG subculture. The model *openai/clip-vit-base-patch16*, combined with a binary classification of swag and no swag, yielded the most consistent and accurate results compared to human ratings in our user study. This outcome was initially unexpected, given the abstract nature of cultural aesthetics and the generalized training dataset used for these models. The implication here is that embedding spaces possess latent semantic associations that allow them to effectively quantify subculturally nuanced aesthetics without explicit fine-tuning for specific cultural constructs.

However, the process also revealed inherent limitations. These models' ability to measure SWAG largely depends on the chosen concept lists, embedding spaces, and the training data used. This highlights that the subjective judgment involved in selecting these concepts, as well as the subjective nature of embedding spaces itself, significantly impacts the outcomes. Therefore, while embeddings can measure culturally subjective aesthetic values, this measurement is heavily contingent on methodological choices and the cultural assumptions embedded in the training data.

The analysis showed cultural biases in both human evaluations and model predictions. Participants rated males higher than females and People of Color higher than Caucasian or Asian individuals. However, some models, like Laion CLIP variants, reversed the gender bias, favoring females. This discrepancy highlights how cultural stereotypes are interpreted differently by humans and models, likely due to biases in the training datasets from the internet.

The observed biases raise significant concerns about cultural representation in machine learning. Latent spaces can both quantify and reinforce existing biases, making it essential to evaluate these cultural patterns carefully when deploying computational models in sensitive contexts.

This research significantly contributes to the transdisciplinary field that connects computer science with culture. By providing an example of how to create a methodological framework that integrates computational tools into cultural analysis, it opens up new possibilities for researchers working at this intersection. The use of SWAG culture serves as an example or placeholder and can be replaced with any other cultural phenomenon or practice that can be sufficiently represented in computational intelligence models.

Our study introduced design artifacts like the Swagometer and Digital Probing Device, which translate complex computational concepts into accessible and interactive formats, improving public engagement and understandability.

Our collaborative approach significantly shaped the project's direction, incorporating our past insights from machine learning, psychology, visual communication, and cultural studies. Workshops and frequent check-ins fostered dynamic exchanges, leading to iterative refinement of research questions and methodologies. Diverse perspectives enriched our analysis but presented challenges in aligning terminologies and methodologies, emphasizing the need for clear communication and structured frameworks.

The employed methodologies, including user studies and interactive visualizations, supported the project's exploratory nature, yet some limitations arose, particularly regarding variability in participant ratings. Future research could improve by utilizing larger-scale studies, robust validation processes, and refined data sampling methods to enhance reliability and generalizability.

7. Conclusion and Future Work

This project presented a speculative methodology for assessing subcultural aesthetics using computational models, framing AI as "Models of Culture." By employing SWAG as a case study, we explored how embedding spaces, typically for classification, can serve as analytical tools for cultural meaning and aesthetic value.

Our findings demonstrated that CLIP-based embedding spaces align surprisingly well with human perceptions of SWAG in binary contexts, though we also noted variability in participant responses and systemic biases in both model predictions and human judgments. This highlights the need for critical examination of training data and methodological assumptions in computational cultural analysis.

From a design perspective, interactive exhibits like the Swagometer translated complex computational concepts into accessible formats, inviting users to engage with the idea of measuring style, a subjective concept.

Future research will refine study design for clearer image interpretation and consistency, while exploring other subcultures to see if similar patterns emerge. We are also interested in expanding our approach to include multi-modal analysis of visual, textual, and audio embeddings.

Ultimately, this research contributes to the discourse on AI and culture, arguing that models trained on cultural media do more than classify—they reflect and reshape culture, making understanding this dynamic crucial for responsible design and analysis in an era of increasing importance of information aesthetics.

8. References

3Blue1Brown. (2024, April 1). *But what is a GPT? visual intro to transformers | chapter 5, deep learning*. Retrieved June 6, 2024, from https://www.youtube.com/watch?v=wjZofJX0v4M

Assmann, J., & Hölscher, T. (Eds.). (1988). *Kultur und gedächtnis* (1. Aufl). Suhrkamp.

Baeyaert, J. (2024, November 5). *The challenge of subjectivity in language models* [Medium]. Retrieved March 15, 2025, from https://medium.com/@joffrey.baeyaert/the-challenge-of-subjectivity-in-language-models-c6553701aaec

Clark, A., & Chalmers, D. J. (1998). The extended mind [Publisher: Oxford University Press]. *Analysis*, *58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Darvill, T. C. (2010). *Concise oxford dictionary of archaeology* (2nd ed). Oxford University Press, Incorporated.

Donald, M. (1993). *Origins of the modern mind: Three stages in the evolution of culture and cognition* (Nachdr.). Harvard University Press.

Engel, D., Hüttenberger, L., & Hamann, B. (2013). A survey of dimension reduction methods for high-dimensional data analysis and visualization (C. Garth, A. Middel, & H. Hagen, Eds.) [Artwork Size: 15 pages, 736056 bytes ISBN: 9783939897460 Medium: application/pdf Publisher: Schloss Dagstuhl – Leibniz-Zentrum für Informatik]. *OASIcs, Volume 27, VLUDS 2011*, *27*, 135–149. https://doi.org/10.4230/OASICS.VLUDS.2011.135

Gm, H., Gourisaria, M. K., Pandey, M., & Rautaray, S. S. (2020). A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review*, *38*, 100285. https://doi.org/10.1016/j.cosrev.2020.100285

Heersmink, R. (2024). History of memory artifacts. In L. M. Bietti & M. Pogacar (Eds.), *The palgrave encyclopedia of memory studies* (pp. 1–12). Springer International Publishing. https://doi.org/10.1007/978-3-030-93789-8_82-1

HeyJodye. (2025, February 8). *Playboi carti's new era: Evolving style & aesthetic breakdown (part 2)| IAMUSIC*. Retrieved March 6, 2025, from https://www.youtube.com/watch?v=jSIC7tpKfz8

Jung, C. G., Franz, M.-L. v., Henderson, J. L., Jacobi, J., & Jaffé, A. (Eds.). (1968). *Man and his symbols*. Dell.

Kroening, K. (2025, February 22). *Most INSANE fashion trends 2025*. Retrieved June 3, 2025, from https://www.youtube.com/watch?v=8wA7vnT3OIw&t=810s

Lamb, A. (2021, February 27). A brief introduction to generative models. https://doi.org/10.48550/arXiv.2103.00265

Levy, O., & Goldberg, Y. (2014, June). Dependency-based word embeddings. In K. Toutanova & H. Wu (Eds.), *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 302–308). Association for Computational Linguistics. https://doi.org/10.3115/v1/P14-2050

Nake, F. (2012). Information aesthetics: An heroic experiment. *Journal of Mathematics and The Arts*, *6*, 65–75. https://doi.org/10.1080/17513472.2012.679458

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021, February 26). Learning transferable visual models from natural language supervision. https://doi.org/10.48550/arXiv.2103.00020

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. https://doi.org/10.1037/h0042519

Ruthotto, L., & Haber, E. (2021). An introduction to deep generative modeling. *GAMM-Mitteilungen*, *44*(2), e202100008. https://doi.org/10.1002/gamm.202100008

Shannon, C. E., & Weaver, W. (1998). *The mathematical theory of communication*. Univ. of Illinois Press.

Simhi, A., & Markovitch, S. (2023, November 9). Interpreting embedding spaces by conceptualization. Retrieved November 7, 2024, from http://arxiv.org/abs/2209.00445

Smith, C. (2014). *Encyclopedia of global archaeology*. Springer Reference.

Underwood, T. (2021, October 21). *Mapping the latent spaces of culture* [The stone and the shell]. Retrieved November 14, 2024, from https://tedunderwood.com/2021/10/21/latent-spaces-of-culture/

Vlachos, M., Domeniconi, C., Gunopulos, D., Kollios, G., & Koudas, N. (2002). Non-linear dimensionality reduction techniques for classification and visualization. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 645–651. https://doi.org/10.1145/775047.775143

Appendix
Study Results

Table A1

Average SWAG Ratings by Image Attributes. This table presents the median SWAG ratings for different image attributes. The first row ("Study") represents the median baseline rating derived from participant data, while subsequent rows correspond to model predictions averaged over all classification combinations. Attributes include gender (male, female), ethnicity (poc, caucasian, asian), age (young, adult), body coverage (full body, upper body), and pose (pose, no pose).

	male	female	poc	caucasian	asian	young	adult	full body	upper body	pose	no pose
Study	44.750	33.750	61.000	31.750	37.500	47.250	29.500	47.250	32.500	32.250	40.250
laion/CLIP-ViT-B-32-laion2B-s34B-b79K	45.446	72.350	68.883	46.962	48.178	48.414	63.122	60.380	51.988	58.982	53.292
laion/CLIP-ViT-bigG-14-laion2B-39B-b160k	46.373	75.095	70.971	53.695	51.731	54.343	65.907	61.940	58.394	63.812	58.416
openai/clip-vit-base-patch16	36.497	47.384	47.448	40.356	36.686	38.874	43.317	41.772	41.207	42.236	39.952
openai/clip-vit-large-patch14-336	31.094	42.134	39.674	37.168	31.531	34.537	39.596	35.292	39.422	39.587	35.791
patrickjohncyh/fashion-clip	49.189	64.595	58.389	49.357	49.595	50.851	59.902	57.361	51.261	58.067	51.065

Table A2

Evaluation Results for All Measurements (Mapping: 1 = swag basic corporate, 2 = swag no swag, 3 = swag sportive style corporate style casual style elegant style unfashionable)

Model	Classif. Combination	Mean MAE	Mean RMSE	Mean p-value	Median MAE	Median RMSE	Median p-value
openai/clip-vit-base-patch16	3	35.348	41.879	0.000	36.903	44.970	0.000
openai/clip-vit-base-patch16	1	37.672	44.512	0.000	39.087	47.208	0.000
openai/clip-vit-base-patch16	2	20.729	24.651	0.304	25.169	29.237	0.547
openai/clip-vit-large-patch14-336	3	37.763	43.930	0.000	39.155	47.002	0.000
openai/clip-vit-large-patch14-336	1	39.266	46.071	0.038	40.795	49.023	0.132
openai/clip-vit-large-patch14-336	2	27.076	33.223	0.000	30.042	36.828	0.000
patrickjohncyh/fashion-clip	3	40.567	46.794	0.019	41.980	49.796	0.010
patrickjohncyh/fashion-clip	1	40.610	47.114	0.002	42.241	50.051	0.012
patrickjohncyh/fashion-clip	2	44.505	50.480	0.000	45.008	52.484	0.000
laion/CLIP-ViT-bigG-14-laion2B-39B-b160k	3	42.179	48.421	0.172	42.564	50.840	0.098
laion/CLIP-ViT-bigG-14-laion2B-39B-b160k	1	41.868	48.238	0.000	42.762	50.675	0.001
laion/CLIP-ViT-bigG-14-laion2B-39B-b160k	2	53.161	57.644	0.000	52.898	59.484	0.000
laion/CLIP-ViT-B-32-laion2B-s34B-b79K	3	42.387	48.709	0.536	43.633	51.614	0.359
laion/CLIP-ViT-B-32-laion2B-s34B-b79K	1	41.884	48.204	0.060	43.190	50.984	0.139
laion/CLIP-ViT-B-32-laion2B-s34B-b79K	2	50.814	55.660	0.000	50.637	57.540	0.000

Table A3

Evaluation Results for High-Confidence Measurements (Mapping: 1 = swag basic corporate, 2 = swag no swag, 3 = swag sportive style corporate style casual style elegant style unfashionable)

Model	Classif. Combination	Mean MAE	Mean RMSE	Mean p-value	Median MAE	Median RMSE	Median p-value
openai/clip-vit-base-patch16	3	40.283	47.232	0.037	43.009	50.766	0.053
openai/clip-vit-base-patch16	1	47.363	55.778	0.005	49.802	58.984	0.008
openai/clip-vit-base-patch16	2	25.767	29.711	0.072	29.736	33.669	0.117
openai/clip-vit-large-patch14-336	3	40.521	48.464	0.019	43.487	51.850	0.031
openai/clip-vit-large-patch14-336	1	47.819	56.292	0.126	50.034	59.199	0.136
openai/clip-vit-large-patch14-336	2	30.904	38.390	0.000	34.468	41.317	0.001
patrickjohncyh/fashion-clip	3	48.830	56.946	0.411	51.296	60.108	0.394
patrickjohncyh/fashion-clip	1	48.396	57.003	0.004	51.072	60.186	0.008
patrickjohncyh/fashion-clip	2	51.726	57.681	0.000	51.754	59.455	0.000
laion/CLIP-ViT-bigG-14-laion2B-39B-b160k	3	47.174	55.676	0.586	49.211	58.590	0.572
laion/CLIP-ViT-bigG-14-laion2B-39B-b160k	1	47.519	56.790	0.004	49.362	59.643	0.006
laion/CLIP-ViT-bigG-14-laion2B-39B-b160k	2	57.993	65.041	0.000	58.150	66.767	0.000
laion/CLIP-ViT-B-32-laion2B-s34B-b79K	3	49.055	58.031	0.108	51.760	61.040	0.129
laion/CLIP-ViT-B-32-laion2B-s34B-b79K	1	49.280	57.338	0.012	51.928	60.447	0.020
laion/CLIP-ViT-B-32-laion2B-s34B-b79K	2	58.583	64.669	0.000	58.254	66.298	0.000