Feeling Textiles through AI: An Exploration into Multimodal Language Models and Human Perception Alignment

Shu Zhong
Department of Computer Science, University College London
London, United Kingdom
shu.zhong.21@ucl.ac.uk

Elia Gatti
Department of Computer Science, University College London
London, United Kingdom
elia.gatti@ucl.ac.uk

Youngjun Cho
Department of Computer Science, University College London
London, United Kingdom
youngjun.cho@ucl.ac.uk

Marianna Obrist
Department of Computer Science, University College London
London, United Kingdom
m.obrist@ucl.ac.uk

Abstract

Human-artificial intelligence (AI) alignment ensures that AI systems align with human goals and behaviors. This paper introduces perceptual alignment as a critical aspect of this alignment, focusing on the concurrence between human judgments and AI evaluations across sensory modalities. We particularly explore how Multimodal Large Language Models (MLLMs), which process both visual and textual data, interpret the tactile qualities of textiles—a significant challenge in online shopping environments. Our research analyzes six vision-based MLLMs to see how they describe the tactile experience of textiles and compares these AI-generated descriptions with human assessments. Through semantic similarity measures and in-person evaluations, we investigate the extent of alignment between human perceptions and AI descriptions. Our findings indicate significant variability in the AI's ability to interpret different textiles, highlighting both the potential and limitations of current AI models in achieving perceptual alignment. This work contributes to understanding the complexities of aligning AI capabilities with human touch sensory experiences.

CCS Concepts

• **Computing methodologies → Artificial intelligence**; • **Human-centered computing**;

Keywords

Human-AI Alignment, Human-AI interaction, Touch Experience, Textile Hand, Multimodal Large Language Models

ACM Reference Format:
Shu Zhong, Elia Gatti, Youngjun Cho, and Marianna Obrist. 2024. Feeling Textiles through AI: An Exploration into Multimodal Language Models and Human Perception Alignment. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '24), November 04–08, 2024, San Jose, Costa Rica.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3678957.3685756

ICMI '24, November 04–08, 2024, San Jose, Costa Rica
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0462-8/24/11
https://doi.org/10.1145/3678957.3685756

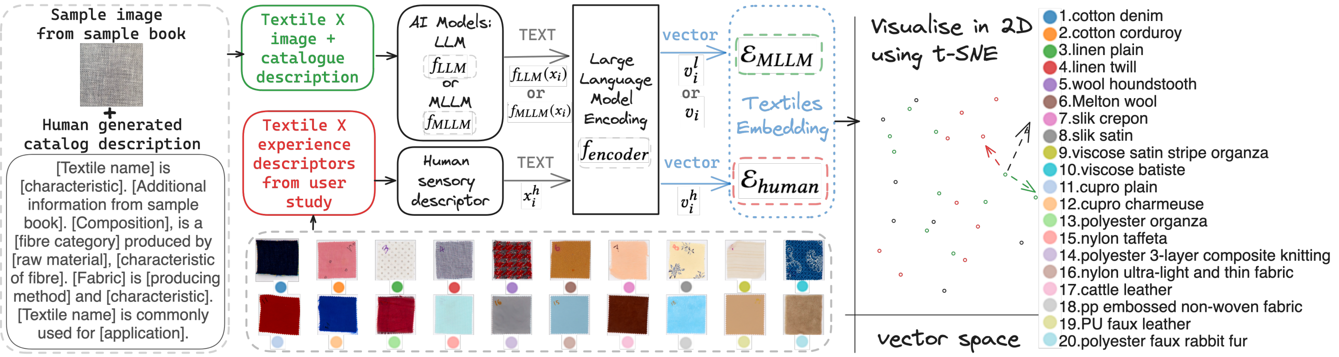
1 Introduction

Artificial Intelligence (AI) lacks human-like perceptions; instead, it processes digitized inputs through algorithms operating on computer systems. For instance, Large Language Models (LLMs), such as GPT-4, mainly handle text-based information, lacking an inherent understanding of touch, smell, or other sensory inputs. On the other hand, Multimodal Large Language Models (MLLMs) like KOSMOS-1 [13] (text+image) process inputs from multiple modalities. They undertake tasks integrating information from various sources such as text and images [13, 15], and potentially tactile data [12, 26].

In the realm of online shopping, the inability to physically touch and feel products often leads to consumer dissatisfaction and high return rates [17]. Can AI understand our experiences of garments, i.e. how they feel like, to help us in daily life? Can current MLLMs bridge human and AI perception by interpreting tactile qualities based on visual and descriptive inputs? Given that vision and text-based descriptions are the primary forms of product presentation in online retail, we focus on vision-based MLLMs.

In this work, we explore how well vision-based MLLMs describe the tactile experiences of textiles compared to human assessments — a concept we refer to as "perceptual alignment". We investigate whether these language models can be well-aligned with humans in assessing the tactile qualities of textiles from textile images and catalog descriptions. Specifically, our study examines the alignment between AI-generated descriptions and human perceptions of textiles' tactile experience. Our study involved an in-person evaluation with 40 participants to compare human and AI interpretations directly. The contributions of this work are as follows:

- We evaluate the ability of six different vision-based MLLMs to interpret the tactile experience of textiles from textiles' appearance and their associated catalog descriptions.
- We compared the generated outputs provided by vision-based MLLMs and by human participants using LLM embeddings to capture more nuanced semantic similarity.
- We explored the variability in AI's ability to understand and interpret different textiles, highlighting the diverse effectiveness of AI across various textile types.



[Textile name] is [characteristic]. [Additional information from sample book]. [Composition], is a [fibre category] produced by [raw material], [characteristic of fibre]. [Fabric] is [producing method] and [characteristic]. [Textile name] is commonly used for [application].

Figure 1: An overview of the proposed method. We use images of textile samples along with their catalog descriptions (leftmost) for Vision-based MLLMs to generate sensory descriptors. Additionally, we present real samples to human participants to elicit descriptors (at the bottom). These descriptors are then input into an LLM encoder to produce textile embeddings ($\mathcal{E}_{MLLM}$ and $\mathcal{E}_{human}$). Subsequently, these embeddings are visualized and analyzed using t-SNE visualization in the vector space (rightmost).

2 Related Work

2.1 Human-AI Alignment

Human-AI alignment refers to the design, development, and refinement of artificial intelligence systems that understand, predict, and augment human intentions and behaviours [10, 23]. Hendrycks et al. presented the ETHICS dataset [10] to evaluate language models' understanding of values – basic moral principles, such as justice and commonsense morality. Further research has built upon this work to reduce toxicity and promote ethical behaviour in language models with human-in-the-loop [9, 21]. Marjieh et al. [18] recently demonstrated that LLMs, i.e. GPT-4, can effectively interpret certain human sensory judgments (e.g., colour, sound and taste) based on textual sensory inputs. For example, they displayed the same pair of colours (red and blue) to both humans (image) and GPT models (hex code), requesting each to rate the similarity score, and then comparing the resulting scores. Their findings show that judgments made by GPT models exhibit correlations with those made by humans. Our research extends beyond these foundational studies by integrating semantic embeddings [4] with MLLMs. Zhong et al. [27] examine LLMs' ability to predict textiles from user touch descriptions but do not compare or categorize human versus AI descriptions, overlooking linguistic nuances and MLLMs.

2.2 Human-AI Alignment in Multimodal LLMs

Multimodal models are particularly powerful in applications where a single modality does not provide enough information to make accurate predictions or decisions. The advent of MLLMs, exemplified by KOSMOS-1 [13], marks a significant advancement in integrating language with perception tasks, such as multimodal dialogue and image recognition with descriptions. Results show that MLLMs can benefit from cross-modal transfer, i.e., transfer knowledge from language to multimodal, and from multimodal to language.

Despite MLLMs exhibiting outstanding performance in multimodal tasks, it is even more crucial to understand their sensory alignment. Extending the exploration of Human-AI alignment in sensory judgments, Lee et al. [15] introduced the VisAlign dataset to evaluate the alignment between AI and human visual perception, aiding in the understanding of aligning AI with human vision perceptual processes. Yet, the sense of touch remains unexplored.

3 Method

We investigate the perceptual alignment between humans and MLLMs in "textile hand". Textile hand refers to the tactile qualities of a textile when touched against the skin [3, 14]. Specifically, we compare the embeddings of their textile hand descriptions for 20 different textiles (see Figure 1 and Section 3.1.1). Marjieh et al. [18] presented a direct method for assessing the perceptual alignment between LLMs and humans, involving human rating of similarity scores. However, this approach overlooks intricacies, such as nuanced semantic similarities. This method can be extended to the embedding space. Consequently, we adopted a novel method that involves encoding human descriptions alongside those generated by various MLLMs about textiles to a high-dimensional embedding space using another LLM encoder to measure the similarity.

3.1 Mapping Textiles to Embeddings

Embeddings are learned representations; they capture the semantics of the input data, group semantically similar items together and keep dissimilar items far apart in the embedding space [11]. LLM encoder assesses similarities by semantic content and overall meaning rather than just lexical. This serves as a valuable measurement tool for our study: if MLLMs are aligned with humans, clearly both MLLM and human-generated outputs would be clustered in proximity when processed by the embedding model. We employ OpenAI's `text-embedding-3-small` [20] to create our embeddings, as this model is among the top-performing on the market.

3.1.1 Data Preparation. We select 20 textile samples based on a combination of a domain-focused taxonomy [28] and textile catalogs [25] that cover a wide range of properties. These 20 textile samples are from four major fiber categories: natural, animal, regenerated, and synthetic, using the TextileNet taxonomy [28]. We chose two widely used materials from each category based on annual consumption [7]. Each sample was sourced from commercial sample books for professionals in design as shown in Figure 1. We then created the textiles' catalog description with domain experts using industry-standard descriptions from Textilepedia [6] and sample books [25]. This ensures the data's relevance for the textile industry and compatibility with current LLMs. Full catalog descriptions are provided in the Appendix.

3.1.2 Map to Embeddings. Previous work has used word embeddings to learn the sensory description language representation in natural language processing [1]. However, traditional word embeddings do not adequately capture the nuanced sentiment and contextual meanings of the entire content as they provide static vectors for each word [19]. This work employs a more sophisticated approach by adopting sentence-level embeddings. Specifically, we use advanced LLM (OpenAI's `text-embedding-3-small` model [20]), which is optimized for representing whole sentences rather than individual word features. This offers context-aware embeddings that adapt to the surrounding text, allowing for a richer and more comprehensive analysis of content sentiment [5, 8].

In other words, each description, generated either by the MLLM ($f_{MLLM}(x_i)$) or human (x_i^h), is encoded by an LLM encoder model ($f_{encoder}$, OpenAI's `text-embedding-3-small`) to generate a unique embedding vector $\{v_i\}$ from the image and text input x_i.

$$v_i = f_{\text{encoder}}(f_{MLLM}(x_i)) \quad (1) \qquad v_i^l = f_{\text{encoder}}(f_{LLM}(x_i)) \quad (2)$$

$$v_i^h = f_{encoder}(x_i^h)) \quad (3)$$

Algorithm 1, 2 and 3 would generate the following sets of vectors: $\mathcal{E}_{MLLM} = \{v_1, v_2, ...v_{20}\}$, $\mathcal{E}_{LLM} = \{v_1^l, v_2^l, ...v_{20}^l\}$ and $\mathcal{E}_{Human} = \{v_1^h, v_2^h, ...v_{20}^h\}$. Taking $\mathcal{E}_{MLLM}$ as an example, Algorithm 1 resulted in 20 generated vectors $\mathcal{E}_{MLLM} = \{v_1, v_2, ...v_{20}\}$, where v_i represents an embedding vector generated by a MLLM. And another 20 vectors for both $\mathcal{E}_{LLM}$ and $\mathcal{E}_{human}$ are generated respectively. This would later enable us to compare the descriptors generated by MLLMs, LLMs and humans. The prompts and experimental settings are detailed in the supplementary material.

For AI's textile description, we consider the current most powerful MLLM/LLM families for both f_{MLLM} and f_{LLM}: OpenAI GPT [20], Google Gemini [24] and Anthropic Glaude [2], and compare the following models via their official APIs:

- OpenAI GPT [20]: `gpt-4` and `gpt-4-turbo-preview`
- Google Gemini [24]: `gemini-1.0-pro` and `gemini-1.5-pro`
- Anthropic Claude [2]: `claude-3-opus-20240229` and `claude-3-sonnet-20240229`

It's worth mentioning that leading LLMs, such as GPT-4, now incorporate the capability to process visual input data, enabling the extraction of both f_{LLM} and f_{MLLM} from the same model. The experiment setup and prompts used are provided in the Appendix.

3.1.3 Visualize Embeddings in 2D Using t-SNE. We run a t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm, which is an unsupervised dimensionality reduction, to visualize LLM embeddings in 2D. This method transforms high-dimensional Euclidean distances between data points into conditional probabilities that reflect their similarities. The probability of point x_i selecting x_j as its neighbor is calculated as follows:

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)} \quad (4)$$

where σ_i represents the variance of the Gaussian distribution centered at point x_i. These probabilities are then symmetrized to incorporate mutual relationships:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \quad (5)$$

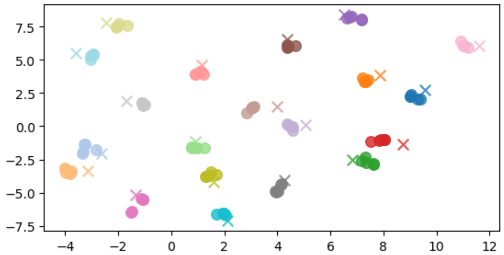


Figure 2: Embeddings of Human created 20 textiles' catalog descriptions *(crosses)* with AI generated descriptions *(dots)* visualized in 2D using t-SNE.

This method effectively maps the similarities onto a 2D plane.

3.2 Human Textile Hand Evaluation

We conducted an in-person study with 40 participants (30 female, 10 male, aged 18-39, mean = 25.79, SD = 4.12) to gather descriptors for 20 selected textile samples. All participants were native or highly proficient English speakers, provided informed consent, and were compensated for their participation. Participants verbally described their 'textile hand' experience after handling textile samples placed inside a black box to avoid bias due to visual cues. Each participant was assigned two textiles to describe and repeat this procedure three times to familiarise them with verbalizing their experiences. Their verbal descriptions were captured using automatic speech recognition (ASR), displayed on a monitor, and confirmed by the participants. These descriptions were then encoded into unique vectors $\mathcal{E}_{human} = \{v_1, ...v_n\}$. This study was approved by the University Research Ethics Committee (Approval ID Number: UCLIC_2021_014_ObristPE).

3.3 Evaluation Measurements

We employ the t-SNE algorithm to visually compare their embeddings in 2D. We first calculate the centroid of the vectors for human descriptions of each textile, represented as $\mathcal{E}_{h_c}$. These centroids capture the average semantic space of human perceptions for each textile. We then compared and visualized these centroids with $\mathcal{E}_{MLLM}$ and $\mathcal{E}_{LLM}$ in the same embedding space using t-SNE. This allows us to observe the clustering and dispersion patterns, comparing how closely AI-generated descriptions resemble human perceptions.

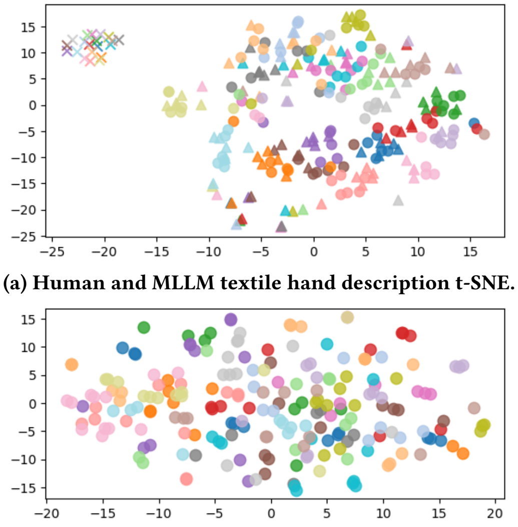
Additionally, we explore the linguistic differences between AI-generated and human descriptions by analyzing term frequency. We first excluded non-substantive words (e.g. "it", "this") and study-specific terms (e.g. "feel", "touch") using Python NLTK. We then used WordCloud for visualisation. This approach emphasizes on the most significant content words used in descriptions, providing insight into the focus and variability of language used.

4 Results and Discussion

We present semantic analysis in embedding space and content analysis for understanding the alignment between AI and humans.

4.1 Alignment with Catalog Descriptions

Our initial experiment does not involve human participants but is centred on investigating the alignment between textile catalog descriptions and descriptions generated by LLMs. The catalog descriptions (see Section 3.1.1) and the corresponding images from the sample books [25]. The aim is to determine if formally defined



(a) Human and MLLM textile hand description t-SNE.

(b) Human textile hand description across 20 textiles t-SNE.

Figure 3: t-SNE for human and AI textile hand descriptions.

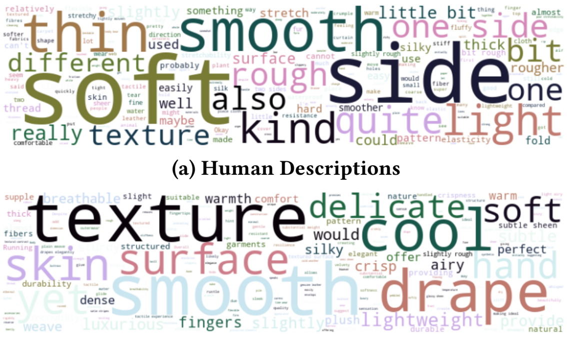
textile descriptions correspond with LLMs' outputs and assess the level of agreement among various LLMs on this specific task.

Figure 2 depicts the t-SNE analysis results for textile catalog descriptions generated by human and different LLMs (f_{LLM}). The proximity of markers in the figure illustrates the similarities between textile descriptions—dots closer together indicate more similarity. Catalog descriptions by humans are marked with crosses, while those generated by LLMs are shown as solid dots. AI-generated catalogs tend to cluster closely, indicating high similarity among them, while human-generated descriptions are distinctly more distanced. Despite these differences, the clustering of catalogs for the same textiles by both humans and LLMs suggests a shared understanding, confirming that both recognize the textiles in a comparable manner.

4.2 Misalignment in Sensory Representation

In this section, our analysis centres on the outcomes of the human textile touch experience detailed in Section 3.2. The experiment specifically targets human descriptors that originate from subjective interpretations, as opposed to the objective catalog descriptions provided in Section 4.1. The result in Figure 3a illustrates participants' textiles hand description *(crosses)* with six MLLMs *(dots)* and six LLMs *(triangles)* textile hand description regarding 20 textiles. Each textile is colour-coded the same way as shown in Figure 1.

We make three interesting observations. First, human descriptions (clustered at the top left in Figure 3a) are distinctly distanced from AI-generated descriptions. This suggests LLM and MLLMs' sensory interpretations are not aligned with humans. Second, the content obtained from humans tends to cluster together, whereas descriptions generated from the AI are more distinguished across different models. Third, it is intriguing to observe that, despite most likely being trained on similar datasets (see Section 3.1.2), the AI models exhibit considerable variance in their sensory descriptions across different textiles, as exemplified by the Cupro Charmeuse (no. 12, light orange) and linen plain (no. 3, green).



(a) Human Descriptions

(b) AI Descriptions

Figure 4: Word clouds from the "textile hand" descriptions.

We then take a closer look at the agreement achieved by participants as shown in Figure 3b. Although human descriptors are quite sparse, humans generally agreed on most of the presented textiles.

4.3 Human vs MLLMs Expressions

The word clouds in Figure 4 present an initial comparison between how MLLMs and humans describe textile hand. At first glance, we observe significant overlap in the vocabulary used, with both clouds featuring prominent terms like "smooth," "texture," and "soft". This indicates a basic alignment in their perception of tactile qualities when characterizing textiles.

However, notable differences also emerge. Humans focus on the immediate sensory insights into what they value in textiles, like practical usage, using terms like "one side," "fold," and "stretchy". Interestingly, some qualities such as "side", "thin", "rough" and "light" are frequently mentioned by humans whereas less by the AI. Additionally, the human descriptions also seem more grounded in personal experience and individual perception, using subjective and qualitative terms like "quite," "kind," and "bit". These terms not only reflect a direct interaction with the textile but also a nuanced understanding of its physical properties and practical functionality.

In contrast, MLLMs offer a more abstract and embellished interpretation, which can sometimes be detached from everyday language. Their language tends to be poetic and emotive, using words such as "drape," "delicate," "luxurious," "perfect," "whisper," and "airy". For instance, `gpt-4-turbo-preview` described viscose satin stripe organza as *"This fabric would caress the skin with a smooth, silky touch, delicately kissed by the subtlest hint of texture from its satin stripes, imparting a luxurious, airy feel with a gentle, fluid drape."* This style contrasts the human tendency to use more tangible, practical terms as P06 *"When I touch it, I can feel the friction, and it's very soft, and the weight of this kind is very light."* AI's rich vocabulary might come from its training on diverse data sources, including marketing language [5]. However, this might lead to a misalignment in interpretations of tactile textile perception, as AI often lacks the human element of subjective sensory experiences.

5 Conclusion and future work

This work is an initial step towards exploring the perceptual alignment between human and AI. Despite some overlap in the vocabulary used to describe 20 textile samples, as evidenced by the alignment with textile catalog descriptions, LLMs' interpretation of how these textiles feel (i.e., textile hand) is not well aligned with humans. While AI excels at processing data and identifying patterns, it often struggles to capture the subjective and nuanced aspects of human perception. These discrepancies highlight the challenge of achieving human-AI perceptual alignment.

To enhance AI's understanding of the physical world, future research needs to integrate more human sensory data into AI training processes, such as Seifi's work on vibration libraries [22]. In addition, our study limited to textiles, exploring textures beyond textile can enrich our understanding of tactile perception. Research should also expand into other daily but underrepresented modalities like olfaction to foster novel sensory interactions, as suggested by Maggioni et al. [16]. Only through a better perceptual alignment, where "better" needs to be carefully understood, can we build future AI systems that understand the physical world and all its multimodal facets that share our everyday experiences.

Acknowledgments

This work was funded by the Engineering and Physical Sciences Research Council (EP/V011766/1) National Interdisciplinary Circular Economy Research (NICER) programme for the UKRI Interdisciplinary Circular Economy Centre for Textiles: Circular Bioeconomy for Textile Materials.

References

[1] Jorge A Alvarado, Carlos Velasco, and Alejandro Salgado. 2023. The organization of semantic associations between senses in language. *Language and Cognition* (2023), 1–30.

[2] AI Anthropic. 2024. Introducing the Next Generation of Claude.

[3] Hassan Behery. 2005. *Effect of mechanical and physical properties on fabric hand.* Elsevier.

[4] Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35, 8 (2013), 1798–1828.

[5] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[6] C. Chan and Fashionary. 2020. *Textilepedia: The Complete Fabric Guide.* Fashionary. https://books.google.co.uk/books?id=d5QRxAEACAAJ

[7] Daniela Coppola. 2021. E-commerce worldwide-Statistics & Facts. *Statista. https://www.statista. com/topics/871/online-shopping/* (2021).

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[9] Nitesh Goyal, Ian D Kivlichan, Rachel Rosen, and Lucy Vasserman. 2022. Is your toxicity my toxicity? exploring the impact of rater identity on toxicity annotation. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–28.

[10] Dan Hendrycks, Collin Burns, Steven Basart, Andrew Critch, Jerry Li, Dawn Song, and Jacob Steinhardt. 2020. Aligning AI With Shared Human Values. In *International Conference on Learning Representations*.

[11] Geoffrey E Hinton and Sam Roweis. 2002. Stochastic neighbor embedding. *Advances in neural information processing systems* 15 (2002).

[12] Yining Hong, Zishuo Zheng, Peihao Chen, Yian Wang, Junyan Li, and Chuang Gan. 2024. MultiPLY: A Multisensory Object-Centric Embodied Large Language Model in 3D World. *arXiv preprint arXiv:2401.08577* (2024).

[13] Shaohan Huang, Li Dong, Wenhui Wang, Yaru Hao, Saksham Singhal, Shuming Ma, Tengchao Lv, Lei Cui, Owais Khan Mohammed, Barun Patra, Qiang Liu, Kriti Aggarwal, Zewen Chi, Johan Bjorck, Vishrav Chaudhary, Subhojit Som, Xia Song, and Furu Wei. 2023. Language Is Not All You Need: Aligning Perception with Language Models. arXiv:2302.14045 [cs]

[14] S Kawabata and Masako Niwa. 1991. Objective measurement of fabric mechanical property and quality: its application to textile and clothing manufacturing. *International Journal of Clothing Science and Technology* 3, 1 (1991), 7–18.

[15] Jiyoung Lee, Seungho Kim, Seunghyun Won, Joonseok Lee, Marzyeh Ghassemi, James Thorne, Jaeseok Choi, O-Kil Kwon, and Edward Choi. 2023. VisAlign: Dataset for Measuring the Alignment between AI and Humans in Visual Perception. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

[16] Emanuela Maggioni, Robert Cobden, Dmitrijs Dmitrenko, Kasper Hornbæk, and Marianna Obrist. 2020. SMELL SPACE: mapping out the olfactory design space for novel interactions. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 5 (2020), 1–26.

[17] Prasenjit Mandal, Preetam Basu, and Kushal Saha. 2021. Forays into omnichannel: An online retailer's strategies for managing product returns. *European Journal of Operational Research* 292, 2 (2021), 633–651.

[18] Raja Marjieh, Ilia Sucholutsky, P v Rijn, Nori Jacoby, and Thomas L Griffiths. 2023. Large language models predict human sensory judgments across six modalities. *arXiv preprint arXiv:2302.01308* (2023).

[19] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[20] Arvind Neelakantan, Tao Xu, Raul Puri, Alec Radford, Jesse Michael Han, Jerry Tworek, Qiming Yuan, Nikolas Tezak, Jong Wook Kim, Chris Hallacy, et al. 2022. Text and code embeddings by contrastive pre-training. *arXiv preprint arXiv:2201.10005* (2022).

[21] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems* 35 (2022), 27730–27744.

[22] Hasti Seifi, Kailun Zhang, and Karon E MacLean. 2015. VibViz: Organizing, visualizing and navigating vibration libraries. In *2015 IEEE World Haptics Conference (WHC)*. IEEE, 254–259.

[23] Ilia Sucholutsky, Lukas Muttenthaler, Adrian Weller, Andi Peng, Andreea Bobu, Been Kim, Bradley C Love, Erin Grant, Jascha Achterberg, Joshua B Tenenbaum, et al. 2023. Getting aligned on representational alignment. *arXiv preprint arXiv:2310.13018* (2023).

[24] Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805* (2023).

[25] ltd.. Texflag Sample Book Co. [n. d.]. Textile Sample Books, fabric sample book. http://www.texflagsample.com/en/

[26] Fengyu Yang, Chao Feng, Ziyang Chen, Hyoungseob Park, Daniel Wang, Yiming Dou, Ziyao Zeng, Xien Chen, Rit Gangopadhyay, Andrew Owens, et al. 2024. Binding Touch to Everything: Learning Unified Multimodal Tactile Representations. *arXiv preprint arXiv:2401.18084* (2024).

[27] Shu Zhong, Elia Gatti, Youngjun Cho, and Marianna Obrist. 2024. Exploring Human-AI Perception Alignment in Sensory Experiences: Do LLMs Understand Textile Hand? *arXiv preprint arXiv:2406.06587* (2024).

[28] Shu Zhong, Miriam Ribul, Youngjun Cho, and Marianna Obrist. 2023. TextileNet: A Material Taxonomy-based Fashion Textile Dataset. *arXiv preprint arXiv:2301.06160* (2023).